

# Dynamic demand for differentiated products with fixed-effects unobserved heterogeneity

VICTOR AGUIRREGABIRIA

University of Toronto, CEPR.

E-mail: [victor.aguirregabiria@utoronto.ca](mailto:victor.aguirregabiria@utoronto.ca)

First version received: 8 May 2022; final version accepted: 13 September 2022.

**Summary:** This paper studies identification and estimation of a dynamic discrete choice model of demand for differentiated product using consumer-level panel data with few purchase events per consumer (i.e., short panel). Consumers are forward-looking and their preferences incorporate two sources of dynamics: *last choice dependence* due to habits and switching costs, and *duration dependence* due to inventory, depreciation, or learning. A key distinguishing feature of the model is that consumer unobserved heterogeneity has a *Fixed Effects* structure; that is, its probability distribution conditional on the initial values of endogenous state variables is unrestricted. I apply and extend recent results to establish the identification of all the structural parameters as long as the dataset includes four or more purchase events per household. The parameters can be estimated using a *sufficient statistic—conditional maximum likelihood* (CML) method. An attractive feature of CML in this model is that the sufficient statistic controls for the forward-looking value of the consumer's decision problem such that the method does not require solving dynamic programming problems or calculating expected present values. **JEL codes:** C23, C25, C51, D12.

**Keywords:** *Structural dynamic discrete choice models, dynamic demand for differentiated products, panel data, fixed effects, habits, switching costs, storable products, durable products.*

## 1. INTRODUCTION

In many markets, consumer demand is dynamic in the sense that consumers' utility depends on past decisions. Sources of dynamics in demand include, among others, storable products,<sup>1</sup> durable

\* I would like to thank helpful comments from the editor, Jaap Abbring, and two anonymous referees. I have benefited from discussions with Jiaying Gu, Matthew Osborne, Jesús Carro, Bo Honoré, Francis Guiton, Lance Lochner, Nail Kashaev, Enrique Sentana, and Orazio Attanasio, and from comments from the audiences in the Applied Economics and Econometrics seminar at Western University, the special session on Econometrics of Dynamic Discrete Choice at the 2021 conference of the Royal Economic Society, and the 2022 conference on econometric methods and empirical analysis of micro data in honour of Manuel Arellano at Banco de España. The price data for Figure 1 in this paper was provided by Nielsen Company through the Kilts Marketing Data Center at The University of Chicago (Nielsen and NielsenIQ, 2022). The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analysing and preparing the results reported herein. This research is supported by funding from the Social Sciences and Humanities Research Council.

<sup>†</sup> Department of Economics, University of Toronto. 150 St. George Street, Toronto, ON, M5S 3G7, Canada.

<sup>1</sup> See Boizot et al. (2001), Pesendorfer (2002), Erdem et al. (2003), and Hendel and Nevo (2006).

products,<sup>2</sup> habit formation and switching costs,<sup>3</sup> adoption costs,<sup>4</sup> and learning.<sup>5</sup> These models exhibit two main forms of *state dependence* in consumers' purchasing decisions: *last choice dependence* and *duration dependence*. We have *last choice dependence* if a consumer's previous purchase of a product has a causal effect on their current probability of buying that product, for instance, as a result of habits, switching costs, or learning. We have *duration dependence* if the time elapsed since the last purchase has a causal effect on the current buying decision. Depletion of storable products and depreciation of durable products generate duration dependence in consumer demand. These forms of *state dependence* can induce substantial differences between short-run and long-run responses of demand to price changes. This has important economic implications in applications such as evaluating the effects of taxes and subsidies, measurement of firms' market power, or consumer welfare. The estimation of dynamic structural demand models using consumer panel data tries to measure these causal effects and use them for counterfactual analysis and welfare evaluation.<sup>6</sup>

Unobserved heterogeneity plays a fundamental role in dynamic demand models using consumer panel data. Ignoring or incorrectly specifying the correlation between unobserved heterogeneity and pre-determined explanatory variables (e.g., previous purchasing decisions) can generate important biases in the estimation of the structural parameters that capture dynamic causal effects (Heckman, 1981b). Furthermore, the distribution of consumer taste heterogeneity has an important impact on demand price elasticities (Berry et al., 1995). The empirical literature on dynamic demand of differentiated product has considered a *Random Effects (RE)* approach to model consumer unobserved heterogeneity. *RE* models impose parametric restrictions on the distribution of unobserved heterogeneity and on the correlation between these unobservables and the initial values of the predetermined explanatory variables.<sup>7</sup> Though this distribution is not identifiable in short panels, its misspecification can generate important biases in the estimation of dynamic causal effects. This is the so called *initial conditions problem* (Heckman, 1981a). Therefore, *RE* models are not robust to misspecification of parametric restrictions on unobserved heterogeneity.<sup>8</sup> *Fixed effects (FE)* approaches impose no restriction on the distribution of consumers' unobserved heterogeneity such that the identification of parameters of interest is more robust than in *RE* models. However, several identification concerns have inclined researchers

<sup>2</sup> See Esteban and Shum (2007), Goettler and Gordon (2011), and Gowrisankaran and Rysman (2012).

<sup>3</sup> See Roy et al. (1996), Keane (1997), and Osborne (2011).

<sup>4</sup> See Ryan and Tucker (2012) and De Groote and Verboven (2019).

<sup>5</sup> See Akerberg (2003) and Ching (2010).

<sup>6</sup> Nevo (2011) and Berry and Haile (2021) are excellent surveys on identification and estimation of static models of demand for differentiated product. For surveys on dynamic models of consumer demand, see Aguirregabiria and Nevo (2013), Keane (2015), and section 6.2 in Gandhi and Nevo (2021).

<sup>7</sup> In the literature of dynamic discrete choice structural models, the most common approach to deal with time-invariant unobserved heterogeneity is assuming that it has a finite mixture distribution with a small number of points of support—often as small as two or three unobserved types. The probability of the initial values of the state variables is unrestricted conditional on an unobserved individual type. Some studies allow the distribution of unobserved types and the probability of initial conditions varies with observable time-invariant individual characteristics. A few studies assume that the probability of the initial conditions comes from the ergodic distribution of the state variables. For more details, see the survey papers by Aguirregabiria and Mira (2010), Arcidiacono and Ellickson (2011), or Keane (2015).

<sup>8</sup> In the context of dynamic discrete choice demand models, a recent paper by Simonov et al. (2020) shows that the bias induced by the misspecification of the probability of the initial conditions can be sizable. The authors show that RE specifications commonly used in empirical applications, if misspecified, can lead to biases larger than two times the true value of the structural parameter. This magnitude of the bias appears even in relatively long panel data sets spanning more than three or four years of shopping history.

to avoid an *FE* approach in applications of dynamic discrete choice structural models, and more specifically, in dynamic demand models.<sup>9</sup>

A first issue is related to the identification of structural parameters. All the existing (positive) identification results in *FE* dynamic discrete choice models impose the restriction that unobserved heterogeneity enters additively in the utility of each choice alternative. However, in dynamic programming models, unobserved heterogeneity enters not only in current utility but also in the continuation value of the forward-looking decision problem, and these continuation values depend nonadditively (and in fact, without a closed-form expression) on both unobserved heterogeneity and observable state variables. The common wisdom was that *FE* models cannot deal with the nonadditive unobserved heterogeneity that is inherent to discrete choice dynamic programming models. A second important issue is related to the fact that *FE* methods cannot deliver identification of the distribution of unobserved heterogeneity with short panels. Most empirical applications of dynamic structural models are interested in using the estimated model to obtain *Average Marginal Effects* (AME) on endogenous variables of changes in explanatory variables or in structural parameters. Demand price elasticities are examples of these AMEs. Until very recently, the common wisdom was that AMEs are not identified in *FE* models, as they are expectations over the distribution of the unobserved heterogeneity (e.g., Abrevaya and Hsu, 2021; Honoré and De Paula, 2021).

Two recent studies provide positive identification results for the two issues discussed above. Aguirregabiria et al. (2021) establish the identification of structural parameters in a class of *FE* dynamic panel data logit models where agents are forward-looking. The class of models includes two types of endogenous state variables: the lagged choice variable, and the time duration in the last choice. Aguirregabiria and Carro (2021) prove point identification of different AMEs in *FE* dynamic logit models. For instance, the average causal effect of changes in the lagged dependent variable or in the duration in last choice are identified.<sup>10</sup>

In this paper, I apply and extend identification results in Aguirregabiria et al. (2021) to an *FE* dynamic panel data model of consumer demand with differentiated products. The model can incorporate storable or durable products, habit formation, and brand switching costs. I present a new identification result in *FE* dynamic discrete choice with forward-looking agents. I show identification of utility parameters associated to state variables that follow exogenous stochastic processes. More specifically, in the context of demand models, I establish the identification of parameters that capture the effect of prices. In an *FE* forward-looking model, this identification result relies on a particular structure in the stochastic process of prices. There are two components in prices: a persistent component and a transitory component (i.e., temporary promotion).

The structural parameters of the model are estimated using a *sufficient statistics—conditional maximum likelihood* (CML) method for dynamic discrete choice panel data models. Seminal work on this method is due to Cox (1958), Rasch (1960, 1961), and Andersen (1970), and in econometrics by Chamberlain (1985). More specifically, I apply the kernel-weighted method in Honoré and Kyriazidou (2000). In the context of dynamic structural models, a very helpful implication of this CML method is that it is not subject to a *curse of dimensionality* associated with the computational cost of solving a dynamic programming problem, or calculating present

<sup>9</sup> An interesting exception is Jones and Landwehr (1988). The authors estimate an *FE* dynamic binary logit model for consumers' decision to buy the same brand as in the last purchase or switching to a different brand. They estimate this model using a conditional maximum likelihood method. In their model, consumers are not forward-looking, there is no duration dependence, there are no product characteristics (e.g., prices), and it is a binary choice model.

<sup>10</sup> In related work, Chernozhukov et al. (2013), and more recently Davezies et al. (2021), provide partial identification results for a general class of AMEs in binary choice panel data models.

values of future utilities. This is particularly relevant in applications of dynamic demand of differentiated product because the dimension of the state space increases exponentially with the number of products, which is typically large. The *sufficient statistics* - CML method "differences out" the continuation (forward-looking) value of the consumer's decision problem, as this value depends on the incidental parameters / unobserved heterogeneity. This implies that continuation values need not be computed to implement the CML estimator. Therefore, the CML estimation of this dynamic structural model is computationally as simple as in static or myopic models, and its cost does not depend on the dimension of the state space.<sup>11</sup>

Related to the curse of dimensionality in the estimation of dynamic structural models, it is worth to note that the CML method in this paper is quite different to the *Finite Dependence—Conditional Choice Probabilities* (FD-CCP) approach in Arcidiacono and Miller (2011). The FD-CCP approach is an RE method that requires parametric assumptions on the distribution of the unobserved heterogeneity, typically, a finite mixture structure. The FD-CCP method also requires consistent nonparametric estimates of conditional choice probabilities for each unobserved individual type. Consistent nonparametric estimates of type-specific CCPs are available—under some conditions—in RE finite mixture models (see Kasahara and Shimotsu, 2009), but not in FE models.

In an influential paper, Hendel and Nevo (2006) propose an approach to deal with the curse of dimensionality in the estimation of structural dynamic discrete choice models of consumer demand. Under the restrictions of (i) no brand switching costs, (ii) no product differentiation in depreciation / depletion rates, and (iii) no product differentiation in consumption (only at the moment of purchase), a consumer brand choice decision is a static decision, and the only dynamic decision is the replenishment or (repurchasing time) decision. This structure—together with the behavioural assumption that consumers believe that McFadden's surplus (i.e., inclusive value) follows a Markov process—substantially facilitates the estimation of the dynamic structural model. In contrast, in this paper, I consider a model that includes brand switching costs, product differentiation in depletion rates, and product differentiation in consumption, so that consumer brand choice is dynamic. Furthermore, the sufficient statistic result in my paper relies on the structure of the utility function, but it does not exploit any additional behavioural restriction on consumers' beliefs.

In a recent paper, Kong et al. (2022) apply the *Dynamic Potential Outcomes* (DPO) model in Torgovitsky (2019) to study the identification of the causal effect of consumers' past purchasing decisions on current choices, and more specifically the separation between brand switching costs and unobserved heterogeneity. The DPO model can be interpreted as the reduced form of a dynamic structural model that allows for a very general form of consumer unobserved heterogeneity including, for instance, unobserved heterogeneity in consumer switching costs. Using the DPO model and consumer panel data—and in the absence of a controlled randomized experiment—the causal effect of state dependence is not point identified. The authors characterized the identified set for the state dependence generated by brand switching costs. They also show that consumer forward-looking behaviour in the presence of brand switching costs implies exclusion restrictions that satisfy the conditions in Abbring and Daljord (2020) to identify the discount factor and the utility function jointly.

<sup>11</sup> CML estimation implies a different type of curse of dimensionality. The evaluation of the conditional likelihood function involves adding up a function over all the possible values of the vector of sufficient statistics. The number of possible values increases exponentially with the number of time periods in the data  $T$ . However, this dimensionality problem has an easy solution—at the cost of some efficiency lost in the CML estimator—which consists of splitting a  $T$ -periods history into shorter sub-histories.

Motivated by the difficulty of estimating RE models with rich forms of unobserved heterogeneity, Chesher and Santos Silva (2002) propose an extension of the standard multinomial logit model and a simple two-step estimation method that shares some similarities with the sufficient-statistics conditional likelihood approach in this paper. Although the authors present their model/method in a static/ cross-sectional context, it seems that it can be applied to a dynamic logit model and combined with the method in this paper to allow for unobserved heterogeneity in the structural parameters capturing state dependence.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 presents identification results. Section 4 describes the estimation method. In Section 5, I summarize and discuss some extensions for further research.

## 2. MODEL

I present a framework that includes both storable and durable differentiated products. Most of the features of the model are standard in the literature. The main distinguishing feature is the *FE* nature of consumer unobserved heterogeneity.

### 2.1. Basics

There are  $J$  different brands of a differentiated product, and we index brands by  $j \in \{1, 2, \dots, J\}$ . We index consumers by  $i$ . Time is discrete and indexed by  $t$ . There is a dichotomy in the definition of *time* in empirical applications of dynamic consumer demand. In most applications,  $t$  has the standard interpretation as calendar time: for instance, the unit of time can be a week. In these applications, the model and data account for time periods where a consumer does not purchase any variety of the differentiated product. In contrast, a good number of empirical applications in this literature consider  $t$  as the index for purchase events, such that  $t = 1$  means a consumer's first purchase,  $t = 2$  is her second purchase, and so on (see Keane, 1997; Osborne, 2011; among others). The different definition of time  $t$  has implications on the interpretation of dynamics in the model, such as duration dependence.<sup>12</sup> Here, I follow the standard definition of  $t$  as calendar time. However, all the identification results in this paper apply also to models where  $t$  indexes purchase events.

Every period  $t$ , consumer  $i$  decides whether to purchase or not one unit of the product, and which brand to buy. Variable  $y_{it} \in \mathcal{Y} = \{0, 1, \dots, J\}$  represents the decision of consumer  $i$  at period  $t$ , where  $y_{it} = 0$  means 'no purchase', and  $y_{it} = j > 0$  means purchase of brand  $j$ . There are two endogenous state variables that capture dynamics in consumer demand: brand choice in the last purchase,  $\ell_{it} \in \{1, 2, \dots, J\}$ ; and time duration since the last purchase,  $d_{it} \in \{1, 2, \dots, \infty\}$ .<sup>13</sup> Last brand choice is related to habit formation and switching costs. Duration since last purchase captures the effect of inventory depletion (for storable products) or depreciation (for durable products). By definition, the transition rule of the vector of endogenous state variables  $\mathbf{x}_{it} \equiv$

<sup>12</sup> Also very importantly, the model where  $t$  indexes purchase events does not include the decision of 'no purchase' as a choice alternative. That is, there is not an *outside alternative*.

<sup>13</sup> At this point, I consider that the space of the duration variable is unbounded from above. In Section 3 below, I introduce Assumption 3.2 that implies that there is a finite value,  $d^*$ , for the duration variable such that a consumer's utility function does not change for values of duration greater than  $d^*$ . Under this Assumption 3.2, we can restrict the space of the duration variable to be  $\{1, 2, \dots, d^*\}$ . Importantly, the value  $d^*$  is identified from the data as long as  $d^* \leq (T - 1)/2$ , where  $T$  is the number of time periods in the data.

$(\ell_{it}, d_{it})$  is:

$$\mathbf{x}_{i,t+1} \equiv (\ell_{i,t+1}, d_{i,t+1}) = f_x(y_{it}, \mathbf{x}_{it}) \equiv \begin{cases} (\ell_{it}, d_{it} + 1) & \text{if } y_{it} = 0 \\ (j, 1) & \text{if } y_{it} = j > 0 \end{cases} \quad (2.1)$$

In many applications, consumers are located in different geographic markets and face different prices. This is not a necessary condition for the identification results in this paper, but I allow for it. Let  $p_{it}(j)$  be the price of product  $j$  at period  $t$  in the market where consumer  $i$  is located, and let  $\mathbf{p}_{it}$  be the vector with the prices of all products,  $\mathbf{p}_{it} \equiv (p_{it}(j) : j = 1, 2, \dots, J)$ .

## 2.2. Utility

Let  $U_{it}$  be the per-period utility that consumer  $i$  obtains at period  $t$ . It has four components:

$$U_{it} = b_i(y_{it}, \ell_{it}, d_{it}) + m_i(y_{it}, \mathbf{p}_{it}) - sc_i(y_{it}, \ell_{it}) + \varepsilon_{it}(y_{it}). \quad (2.2)$$

The term  $b_i(y_{it}, \ell_{it}, d_{it})$  is the utility from consumption of the branded product;  $m_i(y_{it}, \mathbf{p}_{it})$  represents utility from consumption of the outside good or numeraire;  $sc_i(y_{it}, \ell_{it})$  captures brand switching costs or habits; and  $\varepsilon_{it}(y_{it})$  is a consumer idiosyncratic taste shock at the moment of purchase. I describe below each of these components.

**(i) Utility from consumption of the branded product.** The flow utility from consumption of the branded product has the following form:

$$b_i(y_{it}, \ell_{it}, d_{it}) \equiv \begin{cases} \alpha_i(j) & \text{if } y_{it} = j > 0 \\ \alpha_i(\ell_{it}) - \beta^{dep}(\mathbf{w}_i, \ell_{it}) d_{it} & \text{if } y_{it} = 0. \end{cases} \quad (2.3)$$

The model allows for product differentiation at the moment of consumption and not only at the moment of purchase. Parameter  $\alpha_i(j)$  is the flow utility that consumer  $i$  receives from consuming product  $j$  when the product has not depreciated (or depleted) over time. The vector  $\boldsymbol{\alpha}_i \equiv (\alpha_i(1), \alpha_i(2), \dots, \alpha_i(J))$  represents the fixed effects for consumer  $i$ . The fixed effect  $\alpha_i(j)$  depends on product and consumer characteristics in an unrestricted way. For instance, it could have the structure  $\alpha_i(j) = \mathbf{x}'_j \boldsymbol{\beta}_i + \boldsymbol{\xi}'_j \boldsymbol{\omega}_i$ , where  $\mathbf{x}_j$  and  $\boldsymbol{\xi}_j$  are vectors of observable and unobservable product characteristics, respectively, and  $\boldsymbol{\beta}_i$  and  $\boldsymbol{\omega}_i$  are the vectors of marginal utilities of these characteristics for consumer  $i$ . Therefore, the specification of consumer heterogeneity in preferences is very flexible.

The flow utility from consumption of the branded product takes into account depletion or depreciation over time.<sup>14</sup> If the consumer purchases a new product (if  $y_{it} = j > 0$ ), there is no depreciation effect and the flow utility is  $\alpha_i(j)$ . Otherwise, if the consumer does not purchase a new product (if  $y_{it} = 0$ ), they consume product  $\ell_{it}$ —their last purchase—and flow utility is  $\alpha_i(\ell_{it}) - \beta^{dep}(\mathbf{w}_i, \ell_{it}) d_{it}$ , where  $\mathbf{w}_i$  is a vector of consumer characteristics observable to the researcher, e.g., income, education, geographic location. Parameter (or function)  $\beta^{dep}(\mathbf{w}_i, j)$  represents the depreciation (depletion) rate in product  $j$ . This depreciation/depletion effect can vary across products.

**(ii) Utility from consumption of the composite good.** The consumer's budget constraint implies that consumption of the composite good is equal to  $\mu_i - \sum_{j=1}^J p_{it}(j) 1\{y_{it} = j\}$ , where

<sup>14</sup> The datasets used in this literature contain information on consumer purchase histories, but not on consumption and inventories of storable products. To deal with this missing information for storable products, researchers have imposed different restrictions. Aguirregabiria and Nevo (2013) discuss this issue and different restrictions in the literature. A common approach is using duration since last purchase,  $d_{it}$ , as a proxy of inventory. For storable products, this is the approach in Erdem et al. (2003).



where  $\mu_i$  is the consumer's (weekly) disposable income, and  $\sum_{j=1}^J p_{it}(j) 1\{y_{it} = j\}$  represents expenditure in the branded product. The utility from consumption of the composite good is:

$$m_i(y_{it}, \mathbf{p}_{it}) = \gamma(\mathbf{w}_i) h\left(\mu_i - \sum_{j=1}^J p_{it}(j) 1\{y_{it} = j\}\right), \quad (2.4)$$

where  $h(c)$  is a strictly increasing function that is known to the researcher, e.g. linear ( $h(c) = c$ ), logarithmic ( $h(c) = \ln(c)$ ); and  $\gamma(\mathbf{w}_i)$  is a structural parameter (function) related to the marginal utility of consuming the composite good and may depend on consumer observable characteristics  $\mathbf{w}_i$ . When function  $h(c)$  is nonlinear—e.g., logarithmic—the identification results in this paper require that the researcher observes consumer disposable income  $\mu_i$ . For the rest of the paper, I assume that this is the case.

**(iii) Product switching costs / habits.** Following the standard approach, brand switching costs take place at the moment of purchase, and not through the consumption of the product. The specification of brand switching costs is:

$$sc_i(y_{it}, \ell_{it}) = \beta^{sc}(\mathbf{w}_i, \ell_{it}, y_{it}), \quad (2.5)$$

Parameter  $\beta^{sc}(\mathbf{w}_i, k, j)$  represents the cost of switching from brand  $k$  to brand  $j$ . There is no switching cost without switching or without a new purchase, such that  $\beta^{sc}(\mathbf{w}_i, k, k) = 0$  and  $\beta^{sc}(\mathbf{w}_i, k, 0) = 0$ .

This component of the utility can be also interpreted in terms of *habit formation*. In that case, we have that  $sc_i(y_{it}, \ell_{it}) = -\beta^{hab}(\mathbf{w}_i, \ell_{it}, y_{it})$ , where parameter  $\beta^{hab}(\mathbf{w}_i, j, j) \geq 0$  for  $j > 0$  is the increase in utility from purchasing the same brand as in last purchase, and for any  $k \neq j$ ,  $\beta^{hab}(\mathbf{w}_i, k, j)$  is restricted to be zero. This *habit formation* model is equivalent to the *brand switching cost* model under the restrictions  $\beta^{sc}(\mathbf{w}_i, k, j) = \beta^{sc}(\mathbf{w}_i, k', j)$  for any  $k, k' \neq j$ . This more parsimonious specification is quite common in this literature.

Keane (1997, 2015) and Bronnenberg and Dubé (2017) discuss different specifications of brand loyalty/ habit formation. The dependence of utility only on the last purchase is a limited form of habit formation. However, it has been used in a good number of influential studies on brand loyalty, such as Jones and Landwehr (1988), Seetharaman et al. (1999), Shum (2004), Osborne (2011), Pavlidis and Ellickson (2017), Simonov et al. (2020), and Mysliwski et al. (2020), among others. There are simple extensions of this model that provide more flexible forms of habit formation. For instance, the utility of buying product  $j$  may depend on the number of times the product has been chosen during the last  $n$  purchase events, where  $n$  is larger than one. More generally, following the seminal work by Guadagni and Little (1983), the model can include a habits (brand loyalty) stock variable for each product, where the stock depreciates with a certain rate and increases when the product is purchased. Establishing identification of this, more flexible, model of habit formation in this *FE* framework is an interesting topic for further research.

**(iv) Extreme value distributed shocks in preferences.** Finally, variable  $\varepsilon_{it}(j)$  captures other consumer idiosyncratic factors affecting the utility from purchasing product  $j$  at period  $t$ . Variables  $\boldsymbol{\varepsilon}_{it} \equiv (\varepsilon_{it}(0), \dots, \varepsilon_{it}(J))$  are mean-zero i.i.d. over  $(i, t, j)$  with a type I Extreme Value distribution. Both  $\boldsymbol{\alpha}_i$  and  $\boldsymbol{\varepsilon}_{it}$  are unobservable to the econometrician.

Putting together the different components of the utility function, we have:

$$U_{it} = \begin{cases} \alpha_i(\ell_{it}) + \gamma(\mathbf{w}_i) h(\mu_i) - \beta^{dep}(\mathbf{w}_i, \ell_{it}) d_{it} + \varepsilon_{it}(0) & \text{if } y_{it} = 0 \\ \alpha_i(j) + \gamma(\mathbf{w}_i) h(\mu_i - p_{it}(j)) - \beta^{sc}(\mathbf{w}_i, \ell_{it}, j) + \varepsilon_{it}(j) & \text{if } y_{it} = j > 0. \end{cases} \quad (2.6)$$

For the rest of the paper, I use function  $u_{\alpha_i}(y_{it}, \mathbf{x}_{it}, \mathbf{p}_{it})$  to represent the part of the utility that does not include the unobservable logit shocks, such that  $U_{it} = u_{\alpha_i}(y_{it}, \mathbf{x}_{it}, \mathbf{p}_{it}) + \varepsilon_{it}(y_{it})$ .

**(v) A remark on the identification / normalization of switching costs parameters.** In discrete choice models, we can only identify parameters affecting the difference between the utilities of two choice alternatives, for instance, the utility differences  $U_{it}(j) - U_{it}(0)$  for  $j = 1, 2, \dots, J$ . For the model in this paper, this implies that we need some ‘normalization’ restrictions of the switching cost parameters. These restrictions are needed even in the hypothetical (ideal) case where there is no time-invariant consumer unobserved heterogeneity, consumers are not forward-looking, and the researcher observes utility differences  $U_{it}(j) - U_{it}(0)$  for  $j = 1, 2, \dots, J$ . In this model, these utility differences have the following form:

$$U_{it}(j) - U_{it}(0) = \alpha_i(j) - \alpha_i(\ell_{it}) + \gamma [h_{it}(j) - h_{it}(0)] - \beta^{sc}(\ell_{it}, j) + \beta^{dep}(\ell_{it})d_{it} + \varepsilon_{it}(j) - \varepsilon_{it}(0) \quad (2.7)$$

where  $h_{it}(j) \equiv h(\mu_i - p_{it}(j))$  and  $h_{it}(0) \equiv h(\mu_i)$ . To illustrate this issue, consider the simpler case with:  $J = 2$  products;  $\alpha_i(j) = \alpha(j)$  for every consumer  $i$ ; and  $\gamma = 0$ . Conditional on duration  $d_{it} = d$ , we have only four possible utility differences:  $U_{it}(1) - U_{it}(0)$  and  $U_{it}(2) - U_{it}(0)$  for  $\ell_{it} = 1, 2$ . The mean values of these utility differences are:

$$\begin{cases} \mathbb{E}[U_{it}(1) - U_{it}(0)|\ell_{it} = 1, d_{it} = d] = \alpha(1) - \alpha(1) - \beta^{sc}(1, 1) + \beta^{dep}(1) d \\ \mathbb{E}[U_{it}(2) - U_{it}(0)|\ell_{it} = 1, d_{it} = d] = \alpha(2) - \alpha(1) - \beta^{sc}(1, 2) + \beta^{dep}(1) d \\ \mathbb{E}[U_{it}(1) - U_{it}(0)|\ell_{it} = 2, d_{it} = d] = \alpha(1) - \alpha(2) - \beta^{sc}(2, 1) + \beta^{dep}(2) d \\ \mathbb{E}[U_{it}(2) - U_{it}(0)|\ell_{it} = 2, d_{it} = d] = \alpha(2) - \alpha(2) - \beta^{sc}(2, 2) + \beta^{dep}(2) d \end{cases} \quad (2.8)$$

The system of equations in (2.8) includes four restrictions and eight parameters:  $\alpha(1)$ ,  $\alpha(2)$ ,  $\beta^{sc}(1, 1)$ ,  $\beta^{sc}(1, 2)$ ,  $\beta^{sc}(2, 1)$ ,  $\beta^{sc}(2, 2)$ ,  $\beta^{dep}(1)$ , and  $\beta^{dep}(2)$ . We need additional restrictions. As mentioned above, if we interpret  $\beta^{sc}(k, j)$  as a switching cost, it seems natural to assume that there is no switching cost without switching, so that  $\beta^{sc}(1, 1) = \beta^{sc}(2, 2) = 0$ . These additional restrictions imply the identification of the depletion parameters:  $\beta^{dep}(j) d = \mathbb{E}[U_{it}(j) - U_{it}(0)|\ell_{it} = j, d_{it} = d]$ . However, these restrictions are not sufficient to identify the switching cost parameters  $\beta^{sc}(1, 2)$  and  $\beta^{sc}(2, 1)$  separately from the difference in the intercepts  $\alpha(2) - \alpha(1)$ . Nevertheless, it is straightforward to show that we can identify the following linear combination of switching cost parameters:

$$\begin{aligned} & \beta^{sc}(1, 2) + \beta^{sc}(2, 1) - \beta^{sc}(1, 1) - \beta^{sc}(2, 2) = \\ & -\mathbb{E}[U_{it}(2) - U_{it}(0)|\ell_{it} = 1, d_{it} = d] - \mathbb{E}[U_{it}(1) - U_{it}(0)|\ell_{it} = 2, d_{it} = d] \\ & + \mathbb{E}[U_{it}(1) - U_{it}(0)|\ell_{it} = 1, d_{it} = d] + \mathbb{E}[U_{it}(2) - U_{it}(0)|\ell_{it} = 2, d_{it} = d] \end{aligned} \quad (2.9)$$

This result extends to a model with any number of products  $J \geq 2$ . Therefore, for the rest of the paper, I focus on the identification of parameters  $\gamma(\mathbf{w}_i)$ ,  $\beta^{dep}(\mathbf{w}_i, j)$ , and the following linear combination of switching cost parameters:

$$\tilde{\beta}^{sc}(\mathbf{w}_i, k, j) = \beta^{sc}(\mathbf{w}_i, k, j) + \beta^{sc}(\mathbf{w}_i, j, k) - \beta^{sc}(\mathbf{w}_i, k, k) - \beta^{sc}(\mathbf{w}_i, j, j) \quad (2.10)$$

### 2.3. Stochastic process for prices

The structure of the stochastic process of prices that I describe here is not necessary for the identification of the parameters  $\beta^{sc}$  and  $\beta^{dep}$ , but it plays an important role in the identification of the marginal utility of income parameter,  $\gamma$ , in this forward-looking model.



ASSUMPTION 2.1. *The price of any product  $j$  has two components, one that is persistent over time ( $z_{it}(j)$ ), and other that is transitory ( $e_{it}(j)$ ).<sup>15</sup> That is:*

$$p_{it}(j) = \rho(z_{it}(j), e_{it}(j)) \quad (2.11)$$

where  $\rho(\cdot)$  is a function that is strictly increasing in all its arguments, and it is known to the researcher. (A) The vector of persistent components  $\mathbf{z}_{it} \equiv (z_{it}(j) : j = 1, 2, \dots, J)$  follows a first order Markov process with transition density function  $f_z(\mathbf{z}_{i,t+1}|\mathbf{z}_{it})$ . (B) (Conditional independence of the transitory component) Conditional on  $\mathbf{z}_{it}$ , we have that  $\mathbf{z}_{i,t+1}$  and  $\mathbf{e}_{i,t+1}$  do not depend on  $\mathbf{e}_{it}$  or on previous lags of this transitory component. (C) The vectors  $\mathbf{z}_{it} - \mathbf{z}_{i,t-1}$  and  $\mathbf{e}_{it} - \mathbf{e}_{i,t-1}$  have supports that include a neighborhood around zero. ■

Some examples for the  $\rho(\cdot)$  function are the linear function  $p_{it}(j) = z_{it}(j) + e_{it}(j)$ , or the linear in logs function,  $p_{it}(j) = \exp\{z_{it}(j)e_{it}(j)\}$ . See also the example below for the case of Hi-Lo pricing. The stochastic relationship between the prices of the different products is unrestricted.

**EXAMPLE. Hi-Lo pricing.** For many supermarket products, the evolution of weekly prices is characterized by the alternation between a regular price and a promotion price. The (latent) regular and promotion prices remain constant for months, but the occurrence of a promotion is very uncertain and it lasts no more than one week. See, for instance, Hitsch et al. (2019) for recent and extensive evidence using Nielsen data. Figure 1 presents an example of the actual time series of prices for a laundry detergent product.<sup>16</sup> Liu and Balachander (2014) show that this evolution of prices is not consistent with the assumption of first order Markov process that is common in many empirical applications of dynamic demand models. They find that relaxing this restriction has important implications to explain consumer behaviour.

I present here a stochastic process for prices under Hi-Lo pricing that satisfies the general conditions described in Assumption 2.1 above. The persistent component  $z_{it}(j)$  has two elements: the *regular price*,  $z_{it}^{reg}(j)$ , and the *promotion price*,  $z_{it}^{pro}(j)$ . By definition,  $z_{it}^{reg}(j) > z_{it}^{pro}(j)$ . These two (latent) variables have discrete support and follow a joint first order Markov process with substantial persistence.<sup>17</sup> At any period  $t$ , the price of product  $j$  can be either the regular price (i.e.,  $p_{it}(j) = z_{it}^{reg}(j)$ ), or the promotion price (i.e.,  $p_{it}(j) = z_{it}^{pro}(j)$ ). Let  $e_{it}(j)$  be the dummy variable that indicates the existence of a promotion for product  $j$  in market  $i$  at period  $t$ . Therefore, in this example, the  $\rho$  function has the following form:

$$p_{it}(j) = \rho(z_{it}(j), e_{it}(j)) = (1 - e_{it}(j)) z_{it}^{reg}(j) + e_{it}(j) z_{it}^{pro}(j) \quad (2.12)$$

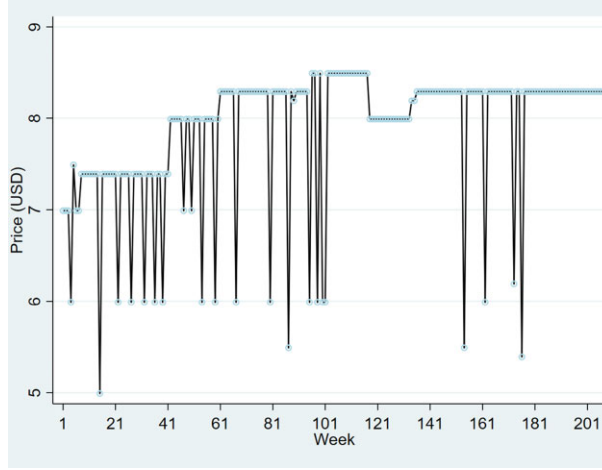
Conditional on  $\mathbf{z}_{it}$ , the vector of promotion dummies  $\mathbf{e}_{it}$  has a multinomial distribution that is independently distributed over time.

Liu and Balachander (2014) consider a stochastic process for prices that satisfies equation (2.12) but where the latent variables for regular and promotion prices are constant over time, and the sales promotion dummy follows a Proportional Hazard Model with duration dependence. Their stochastic process satisfies Assumption 2.1(A) and Assumption 2.1(B) as long as we include the variable ‘*duration since last sales promotion*’ as part of the persistent component,  $\mathbf{z}_{it}$ . However, it does not satisfy Assumption 2.1(C) because the duration variable never remains constant over

<sup>15</sup> Variables  $z_{it}(j)$  and  $e_{it}(j)$  can be scalars or vectors. For instance, in Example 1 below for Hi-Lo pricing,  $z_{it}(j)$  has two components.

<sup>16</sup> The price data for Figure 1 come from the Nielsen Retail Measurement Service scanner dataset, and was provided by Nielsen Company through the Kilts Marketing Data Center at The University of Chicago (Nielsen and NielsenIQ, 2022).

<sup>17</sup> More precisely, the whole vector  $\mathbf{z}_{it}$  for all the  $J$  products follows a first order Markov chain.



**Figure 1.** Weekly price of Tide liquid laundry detergent (70 oz) in a store From week 1-Year 2008 to week 52-Year 2011 (208 weeks).

time, such that the distribution of  $duration_{it} - duration_{i,t-1}$  does not have probability mass around zero. A condition that makes Liu-Balachander's PHM consistent with Assumption 2.1 is that there exists a value of 'duration since last sales promotion', say  $D$ , such that there is no duration dependence either for  $duration_{it} < D$  or for  $duration_{it} > D$ . This restriction is testable. ■

This structure has an important implication on a consumer's dynamic decision model. The whole vector of prices  $\mathbf{p}_t$  (both  $\mathbf{z}_t$  and  $\mathbf{e}_t$ ) affects a consumer's current utility, but the expected and discounted value of future utilities (the continuation value) depends on  $\mathbf{z}_t$  but not on  $\mathbf{e}_t$ . This exclusion restriction plays an important role in the identification of parameter  $\gamma$  in the *FE* dynamic forward-looking model in this paper.

It is relevant to note that, given time series data on prices and a specification of the  $\rho(\cdot)$  function, it is possible to identify the parameters in the two stochastic processes that characterize the evolution of prices and, based on these parameters, it is possible to identify (i.e., to estimate consistently) the two components  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$ . In the example of Hi-Lo pricing above, this identification is straightforward. For the rest of the paper, I assume that these two vectors,  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$ , are observable to the researcher.

#### 2.4. Consumer decision problem

Every period  $t$ , the consumer observes the vectors of state variables  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$ ,  $\mathbf{e}_{it}$ , and  $\mathbf{s}_{it}$ , and makes a purchasing decision  $y_{it}$  to maximize her expected and discounted intertemporal utility  $\mathbb{E}_t \left[ \sum_{s=0}^{\infty} \delta_i^s U_{i,t+s} \right]$ , where  $\delta_i \in [0, 1)$  is consumer  $i$ 's time discount factor.<sup>18</sup> This consumer's problem is a stationary *Markov Decision Process*, and Blackwell's theorem establishes that the value function and the optimal decision rule are time-invariant (Blackwell, 1965).

<sup>18</sup> Consumers know  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$ . For instance, in the example for Hi-Lo pricing above, consumers can distinguish between regular and promotion prices.

The decision problem of consumer  $i$  at period  $t$  is:

$$y_{it} = \arg \max_{j \in \mathcal{Y}} \{u_{\alpha_i}(j, \mathbf{x}_{it}, \mathbf{p}_{it}) + \varepsilon_{it}(j) + v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})\} \quad (2.13)$$

where, as defined in equation (2.1),  $f_x(j, \mathbf{x}_{it})$  represents the value of  $\mathbf{x}_{i,t+1}$  given state  $\mathbf{x}_{it}$  and decision  $y_{it} = j$ ; and  $v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})$  is the *continuation value function*, i.e., the expected and discounted value of future utility given current state is  $(\mathbf{x}_{it}, \mathbf{z}_{it})$  and current choice  $j$ .<sup>19</sup>

A property of the model that plays a key role in the identification of parameter  $\gamma$  is that the continuation value function  $v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})$  does not depend on the transitory component of prices,  $\mathbf{e}_{it}$ . To see this, note that the continuation value is the expectation of next period value function, say  $V_{\alpha_i}(\mathbf{x}_{i,t+1}, \mathbf{z}_{i,t+1}, \mathbf{e}_{i,t+1})$ , conditional on the current value of state variables at period  $t$ ,  $(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it})$ , and on current consumer choice,  $y_{it}$ . To obtain this expectation, one needs to integrate next period value function over the transition density function of the price-related state variables, say  $f_{z,e}(\mathbf{z}_{i,t+1}, \mathbf{e}_{i,t+1} | \mathbf{z}_{it}, \mathbf{e}_{it})$ . The conditional independence in Assumption 2.1(B) implies that this transition density does not depend on  $\mathbf{e}_{it}$ : that is,  $f_{z,e}(\mathbf{z}_{i,t+1}, \mathbf{e}_{i,t+1} | \mathbf{z}_{it}, \mathbf{e}_{it}) = f_{z,e}(\mathbf{z}_{i,t+1}, \mathbf{e}_{i,t+1} | \mathbf{z}_{it})$ . The value of the transitory component of prices at period  $t$  does not contain any information about prices at period  $t + 1$ . This implies that the continuation value function does not depend on  $\mathbf{e}_{it}$ . Note that this property does not depend on any functional form assumption in the utility function, but on the conditional independence in Assumption 2.1(B).

Define the *conditional choice probability* (CCP) function as:

$$P(j | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}, \alpha_i) \equiv Pr \left( j = \arg \max_{k \in \mathcal{Y}} \left\{ u_{\alpha_i}(k, \mathbf{x}_{it}, \mathbf{p}_{it}) + \varepsilon_{it}(k) + v_{\alpha_i}(f_x(k, \mathbf{x}_{it}), \mathbf{z}_{it}) \right\} \middle| \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}, \alpha_i \right) \quad (2.14)$$

This CCP function is conditional on the observable state variables  $(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it})$  and on the unobservable fixed effects  $\alpha_i$ . The type I Extreme Value distribution of the unobservables  $\varepsilon$  implies that the CCP function has the following form. For any  $j \in \mathcal{Y}$ :

$$P(j | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}, \alpha_i) = \frac{\exp \{u_{\alpha_i}(j, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})\}}{\sum_{k=0}^J \exp \{u_{\alpha_i}(k, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(k, \mathbf{x}_{it}), \mathbf{z}_{it})\}} \quad (2.15)$$

The logarithm of this CCP function is:

$$\log P(j | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}, \alpha_i) = u_{\alpha_i}(j, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it}) - \sigma_{\alpha_i}(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) \quad (2.16)$$

where  $\sigma_{\alpha_i}(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it})$  is the logarithm of the denominator in the logit CCP function:

$$\sigma_{\alpha_i}(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) = \log \left( \sum_{k=0}^J \exp \{u_{\alpha_i}(k, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(k, \mathbf{x}_{it}), \mathbf{z}_{it})\} \right) \quad (2.17)$$

### 3. IDENTIFICATION

The researcher observes a panel dataset of  $N$  households over  $T$  periods with information on households' purchasing decisions and prices. The time length of the panel,  $T$ , is short in the

<sup>19</sup> In the literature of structural dynamic discrete models, letter  $v$  is commonly used to denote the *choice-specific value function* or *conditional-choice value function*. Instead, here I use letter  $v$  to denote the continuation value, such that the *choice-specific value function* is  $u_{\alpha_i}(j, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})$ .

sense that it contains only a few purchases per household. The identification results in this paper consider that  $T$  is fixed and—as it is common in proofs of identification—that  $N$  is infinite such that we have an infinite population of households. Also, remember that variables  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$  have been retrieved using data on prices, such that they are observable to the researcher.

Assumption 3.1, together with Assumption 2.1, summarize the restrictions on the model for the identification of structural parameters  $\tilde{\beta}^{sc}$ ,  $\beta^{dep}$ , and  $\gamma$ .

**ASSUMPTION 3.1.** (A) (*i.i.d. Logit shocks*)  $\varepsilon_{it}(j)$  is mean-zero *i.i.d.* over  $(i, t, j)$  with type I Extreme Value distribution, and is independent of  $\alpha_i$ . (B) (*Strict exogeneity of prices with respect to shocks*  $\varepsilon_{it}$ ) For any two periods,  $t$  and  $s$ , the variables  $\varepsilon_{it}(j)$  and prices  $\mathbf{p}_{is}$  are independently distributed. ■

Assumption 3.1(A) and 3.1(B) rule out product-time effects in consumer utility which are common to all consumers in the same local market and unobservable to the researcher. This restriction is very common in the literature of structural dynamic demand models, especially in applications using consumer level data.<sup>20</sup> It is important to note that, until very recently, there was not any point identification result for *FE* dynamic logit models with time dummies (i.e., time fixed effects) using short panels.<sup>21</sup> This may seem a strong assumption in an econometric demand model, and at odds with the literature on estimation of static demand models using the BLP framework (Berry et al., 1995). However, as I explain in the next two paragraphs, the FE approach in this paper controls for two important sources of endogeneity in prices.

First, note that the vector of fixed effects  $\alpha_i = (\alpha_i(1), \alpha_i(2), \dots, \alpha_i(J))$  includes product fixed effects: in fact, it accounts for any interaction of product effects and consumer effects. Therefore, these incidental parameters account for time-invariant differences in product quality that can be correlated with the cross-sectional variation in prices across products. That is, the model accounts—in a very general way—for endogeneity of price levels. This means that the potential concern with Assumption 3.1(B) is because of potential endogeneity of time-variation in prices.

Second, all the identification results in this paper control for variation over time in the persistent component of prices,  $\mathbf{z}_{it}$  and exploit only variation in the transitory component,  $\mathbf{e}_{it}$ . Therefore, the method controls for endogeneity of *regular prices*.

Importantly, we consider an FE model, in the sense that both the unconditional distribution  $F_\alpha(\alpha_i)$  and the sequence of conditional distributions  $F_{\alpha|\mathbf{x},t}(\alpha_i|\ell_{it}, d_{it}, \mathbf{z}_{it}, \mathbf{e}_{it})$  for  $t = 1, 2, \dots, T$  are completely unrestricted functions.

<sup>20</sup> For instance, Erdem et al. (2003), Hendel and Nevo (2006), Pavlidis and Ellickson, (2017), Griffith et al. (2018), or Mysliwski et al. (2020), among many other papers in this literature, assume that prices are not correlated with transitory shocks. In contrast, some dynamic demand models using market-level data, such as Gowrisankaran and Rysman (2012), allow for unobserved market-level demand shocks.

<sup>21</sup> The absence of this identification result is mentioned in Honoré and Kyriazidou (2000), Hahn (2001), and Honoré and Tamer (2006), among many others. Existing FE methods, such as Honoré and Kyriazidou (2000), are based on matching values of exogenous explanatory variables over several time periods. This matching is not possible when the set of explanatory variables includes time dummies, or even a linear time trend. A recent paper by Kitazawa (2022) shows identification of an FE dynamic binary logit model with time dummies and  $T = 4$ . His identification approach uses moment conditions based on an ingenious transformation of the model. Nevertheless, it seems difficult to extend his result to a model where agents are forward-looking because time effects appear also in the agent's continuation value function together with the individual fixed effect.

### 3.1. Sufficient statistics—conditional likelihood approach

Let  $\theta$  be the vector of structural parameters:

$$\theta = (\gamma, \tilde{\beta}^{sc}(k, j), \beta^{dep}(j) : k, j \in \{1, 2, \dots, J\} \text{ with } k > j) \quad (3.1)$$

Note that vector  $\theta$  has finite dimension: more precisely, it has  $1 + J + \frac{J(J-1)}{2}$  elements. To establish the identification of  $\theta$ , I follow Aguirregabiria et al. (2021), who consider a *sufficient statistic—conditional likelihood* approach for dynamic nonlinear panel data models in the spirit of Cox (1958), Rasch (1960), Chamberlain, (1985), or Honoré and Kyriazidou (2000).<sup>22</sup> I start describing some general features of this approach in the context of the dynamic demand model. For notational simplicity, I omit the vector of observable consumer characteristics,  $\mathbf{w}_i$ , for the rest of the paper.

Let  $\mathbf{y}_i \equiv (\ell_{i1}, d_{i1}, y_{i1}, y_{i2}, \dots, y_{iT})$  be the vector with consumer  $i$ 's choice history, including the initial condition  $(\ell_{i1}, d_{i1})$ , and let  $\tilde{\mathbf{z}}_i \equiv (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \dots, \mathbf{z}_{iT})$  and  $\tilde{\mathbf{e}}_i \equiv (\mathbf{e}_{i1}, \mathbf{e}_{i2}, \dots, \mathbf{e}_{iT})$  be the vectors with the histories of the two components in prices. The probability of a choice history  $\mathbf{y}_i$  conditional on the history of prices and parameters  $\theta$  and  $\alpha_i$  is:

$$\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \alpha_i, \theta) = p_1^*(\ell_{i1}, d_{i1} | \alpha_i) \prod_{t=1}^T \frac{\exp \{u_{\alpha_i}(y_{it}, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(y_{it}, \mathbf{x}_{it}), \mathbf{z}_{it})\}}{\sum_{j=0}^J \exp \{u_{\alpha_i}(j, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})\}} \quad (3.2)$$

where  $p_1^*(.)$  is the probability of the state variables at period  $t = 1$ . Given definition of  $\sigma_{\alpha_i}(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it})$ —in equation (2.17) above—as the logarithm of the denominator in the logit CCP function, we have that the log-probability of choice history  $\mathbf{y}_i$  is:

$$\begin{aligned} \log \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \alpha_i, \theta) &= \log p_1^*(\ell_{i1}, d_{i1} | \alpha_i) \\ &+ \sum_{t=1}^T \left[ u_{\alpha_i}(y_{it}, \mathbf{x}_{it}, \mathbf{p}_{it}) + v_{\alpha_i}(f_x(y_{it}, \mathbf{x}_{it}), \mathbf{z}_{it}) - \sigma_{\alpha_i}(\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) \right] \end{aligned} \quad (3.3)$$

Taking into account the structure of the utility function and the transition rules of the state variables, we can develop further this expression to obtain:

$$\begin{aligned} \log \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \alpha_i, \theta) &= \log p_1^*(\ell_{i1}, d_{i1} | \alpha_i) \\ &+ \sum_{t=1}^T 1\{y_{it} = 0\} \left[ \alpha_i(\ell_{it}) + \gamma h_{it}(0) - \beta^{dep}(\ell_{it}) d_{it} \right] \\ &+ \sum_{t=1}^T 1\{y_{it} = 0\} \left[ v_{\alpha_i}(\ell_{it}, d_{it} + 1, \mathbf{z}_{it}) - \sigma_{\alpha_i}(\ell_{it}, d_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) \right] \\ &+ \sum_{t=1}^T \sum_{j=1}^J 1\{y_{it} = j\} \left[ \alpha_i(j) + \gamma h_{it}(j) - \beta^{sc}(\ell_{it}, j) \right] \\ &+ \sum_{t=1}^T \sum_{j=1}^J 1\{y_{it} = j\} \left[ v_{\alpha_i}(j, 1, \mathbf{z}_{it}) - \sigma_{\alpha_i}(\ell_{it}, d_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) \right] \end{aligned} \quad (3.4)$$

where  $1\{.\}$  is the indicator function.

<sup>22</sup> For the autoregressive binary logit model with fixed effects, the derivation of sufficient statistics and the identification of the autoregressive parameter are already established in Cox (1958). See page 227 in that paper.

Based on equation (3.4), the log-probability of a choice history has the following structure:

$$\log \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta}) = \mathbf{s}(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \mathbf{g}(\boldsymbol{\alpha}_i) + \mathbf{c}(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta} \quad (3.5)$$

where  $\mathbf{g}(\boldsymbol{\alpha}_i)$  is a vector of functions of the incidental parameters  $\boldsymbol{\alpha}_i$ , and  $\mathbf{s}_i \equiv \mathbf{s}(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)$  and  $\mathbf{c}_i \equiv \mathbf{c}(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)$  are vectors of statistics, i.e., functions of  $(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)$ . Equation (3.4) provides the elements in each of these vectors. That is:

$$\mathbf{g}(\boldsymbol{\alpha}_i) = \left( \log p_1^*(j, d | \boldsymbol{\alpha}_i), \alpha_i(j), v_{\alpha_i}(j, d, \mathbf{z}), \sigma_{\alpha_i}(j, d, \mathbf{z}, \mathbf{e}) : \right. \\ \left. j = 1, 2, \dots, J; d = 1, 2, \dots, D_T; (\mathbf{z}, \mathbf{e}) \in \mathcal{Z}_T \times \mathcal{E}_T \right) \quad (3.6)$$

where  $D_T$  is the maximum duration observed in the data, and  $\mathcal{Z}_T \times \mathcal{E}_T$  are the sets of values of  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$ , respectively, observed in the data. The following equation presents each element in the vector of statistics  $\mathbf{s}_i$ , with its corresponding element in vector  $\mathbf{g}(\boldsymbol{\alpha}_i)$ .

$$\mathbf{s}_i = \sum_{t=1}^T \begin{bmatrix} 1\{(\ell_{i1}, d_{i1}) = (j, d)\}, \\ 1\{(y_{it}, \ell_{it}) = (0, j)\} + 1\{y_{it} = j\}, \\ 1\{(y_{it}, \mathbf{z}_{it}) = (j, \mathbf{z})\}, \\ 1\{(y_{it}, \ell_{it}, d_{it}, \mathbf{z}_{it}) = (0, j, d-1, \mathbf{z})\}, \\ 1\{(\ell_{it}, d_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) = (j, d, \mathbf{z}, \mathbf{e})\} \end{bmatrix} \rightarrow \begin{bmatrix} \log p_1^*(j, d | \boldsymbol{\alpha}_i), \\ \alpha_i(j), \\ v_{\alpha_i}(j, 1, \mathbf{z}), \\ v_{\alpha_i}(j, d, \mathbf{z}) \text{ for } d \geq 2, \\ \sigma_{\alpha_i}(j, d, \mathbf{z}, \mathbf{e}) \end{bmatrix} \quad (3.7)$$

Similarly, the following equation presents each element in the vector of statistics  $\mathbf{c}_i$ , with its corresponding element in the vector of structural parameters  $\boldsymbol{\theta}$ .

$$\mathbf{c}_i = \sum_{t=1}^T \begin{bmatrix} 1\{(y_{it}, \ell_{it}) = (0, j)\} h_{it}(0) + 1\{y_{it} = j\} h_{it}(j), \\ -1\{(y_{it}, \ell_{it}) = (j, k)\}, \\ -1\{(y_{it}, \ell_{it}) = (0, j)\} d_{it} \end{bmatrix} \rightarrow \begin{bmatrix} \gamma, \\ \tilde{\beta}^{sc}(k, j), \\ \beta^{dep}(j) \end{bmatrix} \quad (3.8)$$

The structure in equation (3.5) has key implications for the identification of  $\boldsymbol{\theta}$ . Suppose that, for every parameter in the vector  $\boldsymbol{\theta}$ , say  $\theta_k$  (the  $k$ -th element of vector  $\boldsymbol{\theta}$ ), there exist two choice histories, say  $A$  and  $B$ , such that  $\mathbf{s}(A) = \mathbf{s}(B)$  and  $\mathbf{c}(A) - \mathbf{c}(B)$  is a vector where all the elements are zero except element  $k$  that is one. Under these conditions, equation (3.5) implies that:

$$\theta_k = \log \mathbb{P}(A) - \log \mathbb{P}(B), \quad (3.9)$$

which shows that parameter  $\theta_k$  is identified from the log odds ratio of histories  $A$  and  $B$ .<sup>23</sup> This intuitive description of the identification of parameters in FE discrete choice models has been used by Chamberlain (1985) and Honoré and Kyriazidou (2000), among others. In this paper, I follow this approach.

More generally, equation (3.5) implies that  $\mathbf{s}_i$  is a *sufficient statistic* for  $\boldsymbol{\alpha}_i$ . That is, the probability  $\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta}, \mathbf{s}_i)$  does not depend on  $\boldsymbol{\alpha}_i$ . To see this, note that  $\mathbf{s}_i$  is a deterministic function of  $\mathbf{y}_i$  such that  $\mathbb{P}(\mathbf{y}_i, \mathbf{s}_i) = \mathbb{P}(\mathbf{y}_i)$ . Then, applying equation (3.5), we have:

$$\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta}, \mathbf{s}_i) = \frac{\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta})}{\mathbb{P}(\mathbf{s}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta})} = \frac{\exp \{ \mathbf{s}_i' \mathbf{g}(\boldsymbol{\alpha}_i) + \mathbf{c}_i' \boldsymbol{\theta} \}}{\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y}) = \mathbf{s}_i} \exp \{ \mathbf{s}_i' \mathbf{g}(\boldsymbol{\alpha}_i) + \mathbf{c}(\mathbf{y})' \boldsymbol{\theta} \}} \\ = \frac{\exp \{ \mathbf{c}_i' \boldsymbol{\theta} \}}{\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y}) = \mathbf{s}_i} \exp \{ \mathbf{c}(\mathbf{y})' \boldsymbol{\theta} \}} = \mathbb{P}(\mathbf{y}_i | \mathbf{s}_i, \mathbf{c}_i, \boldsymbol{\theta}) \quad (3.10)$$

<sup>23</sup> Note that, in this class of models, every choice history has a strictly positive probability such that  $\log \mathbb{P}(A)$  and  $\log \mathbb{P}(B)$  are finite real numbers.



where  $\sum_{\mathbf{y}:s(\mathbf{y})=s_i}$  is the sum over all the possible choice histories  $\mathbf{y}$  with  $s(\mathbf{y})$  equal to  $s_i$ . This sufficiency result comes from usual properties of the exponential family, to which the Type I Extreme Value distribution belongs.

Second, using the expression at the bottom line of equation (3.10), we have the following conditional log-likelihood function (at the population level):

$$\mathbb{E} [\log \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\theta}, s_i)] = \mathbb{E} \left[ \mathbf{c}_i' \boldsymbol{\theta} - \log \left( \sum_{\mathbf{y}:s(\mathbf{y})=s_i} \exp \{ \mathbf{c}(\mathbf{y})' \boldsymbol{\theta} \} \right) \right] \quad (3.11)$$

The first order conditions for the maximization of this likelihood function with respect to  $\boldsymbol{\theta}$  imply the following moment conditions (i.e., likelihood equations):

$$\mathbf{m}(\boldsymbol{\theta}) \equiv \mathbb{E} \left[ \mathbf{c}_i - \sum_{\mathbf{y}:s(\mathbf{y})=s_i} \mathbf{c}(\mathbf{y}) \mathbb{P}(\mathbf{y} | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, \boldsymbol{\theta}, s_i) \right] = \mathbf{0} \quad (3.12)$$

The Jacobian matrix for this vector of moment conditions—or equivalently, the Hessian of the log-likelihood function, or the negative of Fisher's information matrix—is:

$$\frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\mathbb{E} \left( [\mathbf{c}_i - \mathbb{E}(\mathbf{c}_i | s_i)] [\mathbf{c}_i - \mathbb{E}(\mathbf{c}_i | s_i)]' \right) \quad (3.13)$$

This is the negative of a variance-covariance matrix, and therefore it is negative semidefinite for any value of  $\boldsymbol{\theta}$ . In other words, the likelihood function is globally concave in  $\boldsymbol{\theta}$ . Furthermore, based on these moment conditions, a necessary and sufficient condition for (point) identification of  $\boldsymbol{\theta}$  is that this Jacobian matrix  $\partial \mathbf{m}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$  is nonsingular (i.e., rank identification condition). The right-hand-side in equation (3.13) is a nonsingular matrix if there is no perfect collinearity between the residuals of vector  $\mathbf{c}_i$  once we remove its mean dependence on  $s_i$ .

It is important to note that for some FE nonlinear panel data models the identification conditions based on this conditional likelihood approach are sufficient but not necessary. Bonhomme (2012) provides a systematic approach (the *functional differencing* method) to construct moment restrictions for a general class of FE models. Recently, Kitazawa (2013, 2022), Honoré and Weidner (2020), Dobronyi et al. (2021), and Honoré et al. (2021) have used Bonhomme's *functional differencing* approach to establish new point identification results of parameters in FE models where the conditional likelihood approach, at least apparently, does not provide identification.

For the rest of this subsection, I present results on the identification of structural parameters in two versions of the demand model: with and without duration dependence. Here, for simplicity, I do not present the expression for the statistics  $s_i$  and  $\mathbf{c}_i$  and instead provide examples of pairs of histories  $A$  and  $B$  that identify the different parameters in  $\boldsymbol{\theta}$ .

### 3.2. Dynamic demand without duration dependence

Suppose that there is no depreciation or depletion, so that  $\beta^{dep}(j) = 0$  for every product  $j$ , but there is habit formation or/and brand switching costs, so that  $\beta^{sc}(k, j) \neq 0$  for some pair of products  $(k, j)$ . This is the demand model in empirical applications like Roy et al. (1996), Keane (1997), Osborne (2011), or Mysliwski et al. (2020), among others.

In this restricted version of the model,  $\ell_{it}$  is the only state variable whose evolution depends on the consumer's choices. Its transition function is:

$$\ell_{i,t+1} = f_x(y_{it}, \ell_{it}) = \begin{cases} \ell_{it} & \text{if } y_{it} = 0 \\ j & \text{if } y_{it} = j > 0 \end{cases} \quad (3.14)$$

Then, combining equation (2.16) with this transition rule, we have the following expression for the log-CCP function that, for simplicity, I represent using the shorter notation  $\log P_{it}$  instead of  $\log P(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}, \alpha_i)$ :

$$\log P_{it} = \begin{cases} \alpha_i(\ell_{it}) + v_{\alpha_i}(\ell_{it}, \mathbf{z}_{it}) - \sigma_{\alpha_i}(\ell_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) & \text{if } y_{it} = 0 \\ \alpha_i(j) + \gamma h_{it}(j) - \beta^{sc}(\ell_{it}, j) + v_{\alpha_i}(j, \mathbf{z}_{it}) - \sigma_{\alpha_i}(\ell_{it}, \mathbf{z}_{it}, \mathbf{e}_{it}) & \text{if } y_{it} = j > 0 \end{cases} \quad (3.15)$$

where, for notational simplicity, I use  $h_{it}(j)$  to represent  $h(\mu_i - p_{it}(j))$ .

Suppose that  $T = 4$ . Let  $k$  and  $j$  be two different products, and consider the following pair of choice histories:

$$A = (k, j, k, j) \quad ; \quad B = (k, k, j, j) \quad (3.16)$$

Taking into account the structure of the log-CCP function in equation (3.15), we have the following expression for the log-probabilities of the choice histories:

$$\begin{aligned} \log \mathbb{P}(A) = & \log p_1^*(k, \alpha_i) + \alpha_i(j) + \alpha_i(k) + \alpha_i(j) \\ & + v_{\alpha_i}(j, \mathbf{z}_2) + v_{\alpha_i}(k, \mathbf{z}_3) + v_{\alpha_i}(j, \mathbf{z}_4) \\ & - \sigma_{\alpha_i}(k, \mathbf{z}_2, \mathbf{e}_2) - \sigma_{\alpha_i}(j, \mathbf{z}_3, \mathbf{e}_3) - \sigma_{\alpha_i}(k, \mathbf{z}_4, \mathbf{e}_4) \\ & - \beta^{sc}(k, j) - \beta^{sc}(j, k) - \beta^{sc}(k, j) + \gamma (h_2(j) + h_3(k) + h_4(j)), \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} \log \mathbb{P}(B) = & \log p_1^*(k, \alpha_i) + \alpha_i(k) + \alpha_i(j) + \alpha_i(j) \\ & + v_{\alpha_i}(k, \mathbf{z}_2) + v_{\alpha_i}(j, \mathbf{z}_3) + v_{\alpha_i}(j, \mathbf{z}_4) \\ & - \sigma_{\alpha_i}(k, \mathbf{z}_2, \mathbf{e}_2) - \sigma_{\alpha_i}(k, \mathbf{z}_3, \mathbf{e}_3) - \sigma_{\alpha_i}(j, \mathbf{z}_4, \mathbf{e}_4) \\ & - \beta^{sc}(k, j) + \gamma (h_2(k) + h_3(j) + h_4(j)). \end{aligned} \quad (3.18)$$

Therefore, the difference between log-probabilities of the two histories is:

$$\begin{aligned} \log \mathbb{P}(A) - \log \mathbb{P}(B) = & v_{\alpha_i}(j, \mathbf{z}_2) + v_{\alpha_i}(k, \mathbf{z}_3) - v_{\alpha_i}(k, \mathbf{z}_2) - v_{\alpha_i}(j, \mathbf{z}_3) \\ & - \sigma_{\alpha_i}(j, \mathbf{z}_3, \mathbf{e}_3) - \sigma_{\alpha_i}(k, \mathbf{z}_4, \mathbf{e}_4) + \sigma_{\alpha_i}(k, \mathbf{z}_3, \mathbf{e}_3) \\ & + \sigma_{\alpha_i}(j, \mathbf{z}_4, \mathbf{e}_4) \\ & - \tilde{\beta}^{sc}(k, j) + \gamma (h_2(j) - h_3(j) - h_2(k) + h_3(k)). \end{aligned} \quad (3.19)$$

Given equation (3.19), we can establish the identification of parameters  $\tilde{\beta}^{sc}(k, j)$  and  $\gamma$ . To control for the unobserved heterogeneity in the functions  $v(\cdot, \alpha_i)$  and  $\sigma(\cdot, \alpha_i)$ , we need to impose the restrictions  $\mathbf{z}_2 = \mathbf{z}_3 = \mathbf{z}_4$  and  $\mathbf{e}_3 = \mathbf{e}_4$ . Given these restrictions, we have that:

$$\log \mathbb{P}(A) - \log \mathbb{P}(B) = -\tilde{\beta}^{sc}(k, j) + \gamma (h_2(j) - h_3(j) - h_2(k) + h_3(k)). \quad (3.20)$$

Identification of parameters  $\tilde{\beta}^{sc}(k, j)$  and  $\gamma$  follows from equation (3.20). When  $p_2(j) - p_3(j) = 0$  and  $p_2(k) - p_3(k) = 0$ —implying that  $h_2(j) - h_3(j) - h_2(k) + h_3(k) = 0$ —we have identifica-

tion of the switching cost parameter, i.e.,  $\tilde{\beta}^{sc}(k, j) = -\log \mathbb{P}(A) + \log \mathbb{P}(B)$ . Given the identified switching cost and condition  $h_2(j) - h_3(j) - h_2(k) + h_3(k) \neq 0$ , we have identification of the price-sensitivity parameter  $\gamma$ . Remember that equation (3.20) is derived under the condition that the persistent component of prices is constant between periods 2 and 4, and the transitory component is constant between periods 3 and 4. Under these restrictions, we still can have  $h_2(j) - h_3(j) - h_2(k) + h_3(k) \neq 0$  by using variation between periods 2 and 3 in the transitory components of prices of products  $j$  or  $k$ . For instance, in the example of Hi-Lo pricing, we can have a sales promotion for product  $j$  but not for  $k$ , or vice-versa.

This identification result applies to any functional form of  $h(\cdot)$ , linear or nonlinear, as long as the researcher knows this function and observes consumer disposable income. Similarly, this identification result applies to any form of function  $\rho(\cdot)$  in the stochastic process of prices, given that this function is known to the researcher and the transitory component of prices satisfies the conditional independence in Assumption 2.1(B).<sup>24</sup>

The pair of histories  $A$  and  $B$  in equation (3.16) is only one of the many history pairs with identification power for the structural parameters. For instance, we can extend this example by including periods of no purchase. Let  $\mathbf{0}_n$  represent a vector of  $n$  zeros. For  $k, j \geq 1$  with  $k \neq j$ , and any two natural numbers  $n_1$  and  $n_2$ , consider the following choice histories:

$$A = (k, \mathbf{0}_{n_1}, j, \mathbf{0}_{n_2}, k, \mathbf{0}_{n_2}, j) \quad ; \quad B = (k, \mathbf{0}_{n_1}, k, \mathbf{0}_{n_2}, j, \mathbf{0}_{n_2}, j) \quad (3.21)$$

It is straightforward to show that the difference between the log-probabilities of these two histories has the following expression:

$$\begin{aligned} \log \mathbb{P}(A) - \log \mathbb{P}(B) &= \sum_{t=n_1+2}^{n_1+n_2+2} v_{\alpha_i}(j, \mathbf{z}_t) - v_{\alpha_i}(k, \mathbf{z}_t) + \sum_{t=n_1+n_2+3}^{n_1+2n_2+3} v_{\alpha_i}(k, \mathbf{z}_t) - v_{\alpha_i}(j, \mathbf{z}_t) \\ &\quad - \sum_{t=n_1+3}^{n_1+n_2+3} \sigma_{\alpha_i}(j, \mathbf{z}_t, \mathbf{e}_t) - \sigma_{\alpha_i}(k, \mathbf{z}_t, \mathbf{e}_t) \\ &\quad + \sum_{t=n_1+n_2+4}^{n_1+2n_2+4} \sigma_{\alpha_i}(k, \mathbf{z}_t, \mathbf{e}_t) - \sigma_{\alpha_i}(j, \mathbf{z}_t, \mathbf{e}_t) \\ &\quad - \tilde{\beta}^{sc}(k, j) + \gamma (h_{n_1+2}(j) - h_{n_1+n_2+3}(j) - h_{n_1+2}(k) + h_{n_1+n_2+3}(k)). \end{aligned}$$

To eliminate the unobserved heterogeneity  $\alpha_i$  from this difference we need the following conditions on prices: (i) the permanent component  $\mathbf{z}_t$  is constant from period  $n_1 + 2$  to  $n_1 + 2n_2 + 4$ ; and (ii) the transitory component  $\mathbf{e}_t$  is constant between periods  $n_1 + 3$  and  $n_1 + 2n_2 + 4$ . Under these conditions, we have that:

$$\log \mathbb{P}(A) - \log \mathbb{P}(B) = -\tilde{\beta}^{sc}(k, j) + \gamma (h_{n_1+2}(j) - h_{n_1+3}(j) - h_{n_1+2}(k) + h_{n_1+3}(k)). \quad (3.23)$$

This equation shows that a change between periods  $n_1 + 2$  and  $n_1 + 3$  in the transitory component of the price of product  $j$  or  $k$  identifies parameter  $\gamma$ . The switching cost parameter  $\tilde{\beta}^{sc}(k, j)$  is identified from histories where this transitory component is constant.

<sup>24</sup> On a minor technical point, remember that both  $h(\cdot)$  and  $\rho(\cdot)$  are strictly monotonic functions such that  $h_{it}(j) \equiv h(\mu_i - \rho[z_{it}(j), e_{it}(j)])$  always varies when  $e_{it}(j)$  does.

### 3.3. Dynamic demand with duration dependence

Consider now the demand model with duration dependence. The expression for the log-CCP function is:

$$\log P_{it} = \begin{cases} \alpha_i(\ell_{it}) - \beta^{dep}(\ell_{it})d_{it} + v_{\alpha_i}(\ell_{it}, d_{it} + 1, \mathbf{z}_t) - \sigma_{\alpha_i}(\ell_{it}, d_{it}, \mathbf{z}_t, \mathbf{e}_t) & \text{if } y_{it} = 0 \\ \alpha_i(j) + \gamma h_{it}(j) - \beta^{sc}(\ell_{it}, j) + v_{\alpha_i}(j, 1, \mathbf{z}_t) - \sigma_{\alpha_i}(\ell_{it}, d_{it}, \mathbf{z}_t, \mathbf{e}_t) & \text{if } y_{it} = j > 0 \end{cases} \quad (3.24)$$

First, it is straightforward to verify that the pair of histories in equation (3.16) still identifies the switching cost parameters  $\beta_{kj}^{sc}$  and the price-sensitivity parameter  $\gamma$ . Therefore, I focus here on the identification of the depreciation parameters  $\beta^{dep}(j)$ . Furthermore, for notational simplicity, I consider here that the two price components are constant over the considered choice histories and omit  $\mathbf{z}_t$  and  $\mathbf{e}_t$  as arguments.

Let  $n$  be a natural number such that  $2 \leq n \leq (T - 2)/2$ . Consider the following choice histories, both with initial duration  $d_1 = 1$ :

$$A = (j, \mathbf{0}_{n-1}, j, \mathbf{0}_{n+1}) \quad ; \quad B = (j, \mathbf{0}_n, j, \mathbf{0}_n) \quad (3.25)$$

Note that these histories do not contain any product switching event such that switching cost parameters do not appear in the probabilities of these choice histories. Taking into account the structure of the log-CCP in equation (3.24), we have the following expressions for log-probabilities of these choice histories:

$$\begin{aligned} \log \mathbb{P}(A) = & \log p_1^*(j, d_1 | \alpha_i) + 2(n+1)\alpha_i(j) - \sum_{d=1}^{n-1} \beta^{dep}(j)d - \sum_{d=1}^{n+1} \beta^{dep}(j)d \\ & + \sum_{d=2}^n v_{\alpha_i}(j, d) + \sum_{d=1}^{n+2} v_{\alpha_i}(j, d) - \sum_{d=1}^n \sigma_{\alpha_i}(j, d) - \sum_{d=1}^{n+1} \sigma_{\alpha_i}(j, d). \end{aligned} \quad (3.26)$$

$$\begin{aligned} \log \mathbb{P}(B) = & \log p_1^*(j, d_1 | \alpha_i) + 2(n+1)\alpha_i(j) - \sum_{d=1}^n \beta^{dep}(j)d - \sum_{d=1}^{n-1} \beta^{dep}(j)d \\ & + \sum_{d=2}^{n+1} v_{\alpha_i}(j, d) + \sum_{d=1}^{n+1} v_{\alpha_i}(j, d) - \sum_{d=1}^{n+1} \sigma_{\alpha_i}(j, d) - \sum_{d=1}^n \sigma_{\alpha_i}(j, d). \end{aligned} \quad (3.27)$$

They imply the following expression for the difference between the log-probabilities:

$$\log \mathbb{P}(A) - \log \mathbb{P}(B) = -\beta^{dep}(j) + v_{\alpha_i}(j, n+2) - v_{\alpha_i}(j, n+1) \quad (3.28)$$

The expression in the right-hand-side still depends on the incidental parameters  $\alpha_i$  such that, without further restrictions, this pair of histories does not identify parameter  $\beta^{dep}(j)$ .

In general, without further restrictions, the depreciation rates  $\beta^{dep}(j)$  are not identified in the forward-looking model.<sup>25</sup> To obtain identification of these parameters, I follow the same approach as Aguirregabiria et al. (2021) and impose the following restriction.

<sup>25</sup> Note that in a myopic model (i.e.,  $\delta_i = 0$ ) all the continuation values  $v_{\alpha_i}(j, d)$  are zero, such that Equation 3.28 implies the identification of  $\beta^{dep}(j)$  from  $\log \mathbb{P}(A) - \log \mathbb{P}(B)$ . Here, I do not impose this restriction and consider that consumers can be forward-looking.

ASSUMPTION 3.2. For any product  $j$ , there is a value of duration  $d_j^*$ —which can vary across products—such that the structural function that captures the depreciation effect is:

$$\begin{cases} \beta^{dep}(j)n & \text{if } n \leq d_j^* \\ \beta^{dep}(j)d_j^* & \text{if } n \geq d_j^*. \end{cases} \quad \blacksquare \quad (3.29)$$

Importantly, as established in proposition 6 in Aguirregabiria et al. (2021), the value of  $d_j^*$  is identified from the data as long as it is not larger than  $(T - 1)/2$ . I reproduce here this result from Aguirregabiria et al. (2021).

**PROPOSITION.** Let  $d_j^*$  be the value of duration defined in Assumption 3.2 above. For any product  $j$  and any duration  $n$  with  $2n + 1 \leq T$ , define the pair of histories  $A_{j,n} = (j, \mathbf{0}_{n-1}, j, \mathbf{0}_{n+1})$  and  $B_{j,n} = (j, \mathbf{0}_n, j, \mathbf{0}_n)$ . If  $d_j^* \leq (T - 1)/2$ , then the value  $d_j^*$  is point identified from the following expression:

$$d_j^* = \max\{n : \log \mathbb{P}(A_{j,n}) - \log \mathbb{P}(B_{j,n}) \neq 0\} \quad \blacksquare \quad (3.30)$$

An important implication of Assumption 3.2 is that the continuation value function is such that  $v_{\alpha_i}(j, n) = v_{\alpha_i}(j, d_j^*)$  for any duration  $n \geq d_j^*$ . Combining this property with equation (3.28), we have that for  $n = d_j^* - 1$ :

$$\log \mathbb{P}(A_{j,n}) - \log \mathbb{P}(B_{j,n}) = -\beta^{dep}(j) \quad (3.31)$$

such that the depreciation rate  $\beta^{dep}(j)$  is identified.

## 4. ESTIMATION

### 4.1. Estimation of structural parameters

Let  $\theta$  be the vector of structural parameters as defined in equation (3.1). The dataset is  $\{y_{it}, \mathbf{z}_{it}, \mathbf{e}_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ . Let  $\tilde{\mathbf{z}}_i$  and  $\tilde{\mathbf{e}}_i$  the vectors with the time series of prices  $\{y_{it} : t = 1, 2, \dots, T\}$  and  $\{\mathbf{e}_{it} : t = 1, 2, \dots, T\}$ , respectively.

In the identification results in Section 3, we represent a sufficient statistic for  $\alpha_i$  as a binary indicator that combines two conditions: the condition that choice history  $\mathbf{y}_i$  is either equal to  $A$  or to  $B$ ; and a set of restrictions on the evolution of prices. For instance, for the identification of switching costs and price parameters, the persistent component  $\mathbf{z}_{it}$  should be constant between periods  $n_1 + 2$  and  $n_1 + 2n_2 + 4$ , and the transitory component  $\mathbf{e}_{it}$  should be constant between periods  $n_1 + 3$  and  $n_1 + 2n_2 + 4$ . We represent these restrictions on prices using the system of equations  $r(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}$ . Therefore, a sufficient statistic  $s_i$  is the following binary indicator:

$$s_i \equiv s(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = 1\{\mathbf{y}_i \in A \cup B \text{ and } r(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}\} \quad (4.1)$$

We have shown that:

$$\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i, s_i = 1) = \frac{\exp\{c(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \theta\}}{\exp\{c(A, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \theta\} + \exp\{c(B, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \theta\}} \quad (4.2)$$

where  $c(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)$  is a vector of known statistics.

There are many pairs of histories  $A$  and  $B$  that provide sufficient statistics for  $\alpha_i$  and have identification power for  $\theta$ . Let these sufficient statistics be indexed by  $m \in \{1, 2, \dots, M\}$ . We also index by  $m$  the different elements that define sufficient statistic  $s_i^m$ , that is: the corresponding

pair of choice histories,  $(A^m, B^m)$ ; the restrictions on prices,  $r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}$ ; and the identifying statistics,  $c^m(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)$ . Note that we can use  $m$  also to index different sub-periods in the panel dataset.<sup>26</sup> Given these  $M$  sufficient statistics, we can define the following *conditional log-likelihood function*:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = & \sum_{m=1}^M \sum_{i=1}^N 1\{\mathbf{y}_i \in A^m \cup B^m\} 1\{r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}\} \\ & \times \log \left( \frac{\exp\{c^m(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\}}{\exp\{c^m(A^m, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\} + \exp\{c^m(B^m, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\}} \right) \end{aligned} \quad (4.3)$$

The CML estimator is the value of  $\boldsymbol{\theta}$  that maximizes  $\mathcal{L}(\boldsymbol{\theta})$ , which is a globally concave function.

In many possible applications, such as those using weekly or daily supermarket scanner data, price stickiness implies such that the restrictions  $r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}$  hold for a nonnegligible fraction of observations. Nevertheless, imposing exactly these restrictions typically implies losing a substantial amount of observations. This is exactly the issue that motivates the kernel-weighted CML in Honoré and Kyriazidou (2000). In the log-likelihood function, we replace the indicator  $1\{r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i) = \mathbf{0}\}$  with a weight that depends inversely on the magnitude of vector  $|r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)|$  such that we put more weight in the log-likelihood of observations for which  $|r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)|$  is close to zero. More specifically, we consider the following kernel-weighted conditional log-likelihood function:

$$\begin{aligned} \mathcal{L}^{KW}(\boldsymbol{\theta}) = & \sum_{m=1}^M \sum_{i=1}^N 1\{\mathbf{y}_i \in A^m \cup B^m\} K \left( \frac{r^m(\tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)}{b_N} \right) \\ & \times \log \left( \frac{\exp\{c^m(\mathbf{y}_i, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\}}{\exp\{c^m(A^m, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\} + \exp\{c^m(B^m, \tilde{\mathbf{z}}_i, \tilde{\mathbf{e}}_i)' \boldsymbol{\theta}\}} \right) \end{aligned} \quad (4.4)$$

where  $K(\cdot)$  is a kernel density function that satisfies the regularity condition  $K(v) \rightarrow 0$  as  $\|v\| \rightarrow \infty$ , and  $b_N$  is a bandwidth parameter such that  $b_N \rightarrow 0$  and  $Nb_N \rightarrow \infty$  and  $N \rightarrow \infty$ . As shown by Honoré and Kyriazidou (2000), this estimator is consistent and asymptotically normal. If prices  $\mathbf{z}_{it}$  and  $\mathbf{e}_{it}$  have discrete support, the rate of convergence of the estimator is  $N^{-1/2}$ . Otherwise, with continuous prices, the rate of convergence is slower than  $N^{-1/2}$ .

#### 4.2. Average marginal effects and counterfactual experiments for aggregate consumer demand

An important motivation for the estimation of structural demand models is using them to obtain *Average Marginal Effects* (AMEs) and *counterfactual experiments* that provide the causal effect on aggregate consumer demand of hypothetical changes in structural parameters or in state variables. In the context of dynamic demand models, short-run and long-run aggregate demand elasticities are important examples of AMEs. Possible counterfactual experiments include evaluating the effect on aggregate consumer demand of removing sources of dynamics (e.g., making  $\tilde{\beta}^{sc}$  or  $\beta^{dep}$  equal to zero), or introducing hypothetical new products.<sup>27</sup>

In FE discrete choice models with short panels, the distribution of the time-invariant unobserved heterogeneity is not identified. This is because the data consist of a finite number of probabilities—

<sup>26</sup> For instance, we can divide a panel dataset with  $T = 12$  periods into three different sub-panels with four time periods each, and construct sufficient statistics separately for each sub-panel.

<sup>27</sup> Some of these counterfactuals may require endogenizing firms' pricing decisions, but, as a first step, they can be implemented keeping prices fixed.



as many as the number of possible choice histories—but the distribution of the unobserved heterogeneity has infinite dimension. Since aggregate consumer demand is an expectation over the distribution of the unobserved heterogeneity, and this distribution is not identified, the common wisdom is that FE approaches cannot (point) identify AMEs or counterfactuals on aggregate demand. Here I present a brief discussion of two possible approaches.

**(i) Random Effects approach.** After the FE estimation of the structural parameters  $\theta$ , the researcher can consider a parametric RE model for the distribution of the unobserved heterogeneity,  $f_{\alpha}(\alpha_i)$ , and the probability function for the initial conditions,  $p^*(\ell_{i1}, d_{i1}|\alpha_i)$ . The parameters in these two functions can be estimated using full maximum likelihood (or other full solution method), taking as given the estimate of  $\theta$ . Given the FE estimate of  $\theta$  and the RE estimate of the parameters in  $f_{\alpha}(\alpha_i)$  and  $p^*(\ell_{i1}, d_{i1}|\alpha_i)$ , the researcher can calculate any AME or counterfactual experiment.

Though conceptually simple, this approach suffers from two important limitations. A first issue is the potential misspecification of the parametric restrictions imposed on functions  $f_{\alpha}(\alpha_i)$  and  $p^*(\ell_{i1}, d_{i1}|\alpha_i)$ . Though the FE estimation of  $\theta$  is robust to this misspecification, the estimates of AMEs and counterfactuals are not, as they depend on the distribution of the unobservables. A second issue is that the full solution estimation of the model requires computing continuation value functions that may be impractical in applications when the number of products is not small.

**(ii) Recent results on identification of AMEs in FE discrete choice models.** Here I briefly describe recent developments on identification and estimation of AMEs and counterfactuals in FE discrete choice models with short panels. In contrast to the RE approach in item (i) above, these methods share the same attractive features of robustness and computational simplicity in the FE estimation of  $\theta$ . However, this literature is very recent and the results are still limited.

Chernozhukov et al. (2013) study the identification and estimation of AMEs and counterfactuals in FE binary choice models. The AMEs of interest consist of the causal effect on aggregate demand (i.e., CCP function integrated over the distribution of unobserved heterogeneity) of a change in an explanatory variable for all the consumers (e.g., a change in market prices) or a change in structural parameters (e.g., making  $\beta^{dep}$  equal to zero). The authors characterize the identified set of the AME and propose an algorithm to estimate this set.

For FE dynamic multinomial logit models as the ones in this paper, Aguirregabiria and Carro (2021) show the point identification of AMEs that consist of a change in the dynamic explanatory variables, i.e., a change in product choice in last purchased ( $\ell_{it}$ ), or in duration since last purchased ( $d_{it}$ ). Their identification results apply to the model in this paper. The authors show that these AMEs are relatively simple expressions of the probability of some choice histories and of structural parameters. In a similar spirit, Davezies et al. (2021) and Pakel and Weidner (2021) present new results on set identification of a broader class of AMEs in FE discrete choice models. They also propose simple algorithms to compute the identified sets.

## 5. CONCLUSIONS

This paper presents an FE dynamic panel data model of demand for different products where consumers are forward-looking. I apply and extend recent results from Aguirregabiria et al. (2021) to establish the identification of all structural parameters in this model. Several extensions of the results in this paper are interesting topics for further research.

First, the specification of the sources of demand dynamics in the model of this paper is restrictive. For instance, as mentioned in Section 2.2, more realistic models of habit formation

include habit stock variables for each product, where the stock increases when the product is purchased and declines otherwise. Also, the model does not accommodate some forms of consumer learning for experienced goods used in the literature with RE models (e.g., Ching, 2010).

Second, the model assumes that consumers buy at most one unit of the product per period. However, it is well known that forward-looking consumers can buy for inventory (Hendel and Nevo, 2006, 2013). It would be interesting to extend the model to incorporate the possibility of consumers purchasing multiple units.

Third, as mentioned in Section 4.2, an important motivation for the estimation of structural demand models is the evaluation of counterfactual experiments on aggregate consumer demand. Developments on (point or set) identification and estimation of this type of counterfactuals would be very useful in this literature.

Fourth, in the same spirit as the recent work by Mysliwski et al. (2020), this dynamic demand can be combined with a dynamic game of price competition. Assumption 2.1 on the two price components—persistent and transitory—is consistent with an equilibrium of a dynamic pricing game with i.i.d. firms' private information.

## REFERENCES

- Abbring, J. H. and Ø. Daljord (2020). Identifying the discount factor in dynamic discrete choice models. *Quantitative Economics* 11(2), 471–501.
- Abrevaya, J. and Y. Hsu (2021). Partial effects in nonlinear panel data models with correlated random effects. *The Econometrics Journal* 24(3), 519–35.
- Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination\*. *International Economic Review* 44(3), 1007–40.
- Aguirregabiria, V. and J. Carro (2021). Identification of average marginal effects in fixed effects dynamic discrete choice models. *CEPR Discussion Paper No. DP16354*.
- Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: a survey. *Journal of Econometrics* 156(1), 38–67.
- Aguirregabiria, V. and A. Nevo (2013). Recent developments in empirical IO: dynamic demand and dynamic games. *Advances in Economics and Econometrics* 3, 53–122.
- Aguirregabiria, V., J. Gu and Y. Luo (2021). Sufficient statistics for unobserved heterogeneity in dynamic structural logit models. *Journal of Econometrics* 223(2), 280–311.
- Andersen, E. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32, 283–301.
- Arcidiacono, P. and P. B. Ellickson (2011). Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics* 3(1), 363–94.
- Arcidiacono, P. and R. A. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79(6), 1823–67.
- Berry, S. T. and P. A. Haile (2021). Foundations of demand estimation. In Kate Ho, Ali Hortaçsu and Alessandro Lizzeri (Eds.), *Handbook of Industrial Organization, Volume 4*, 1–62. Elsevier.
- Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–90.
- Blackwell, D. (1965). Discounted dynamic programming. *Annals of Mathematical Statistics* 36(1), 226–35.
- Boizot, C., J.-M. Robin and M. Visser (2001). The demand for food products: an analysis of interpurchase times and purchased quantities. *Economic Journal* 111(470), 391–419.

- Bonhomme, S. (2012). Functional differencing. *Econometrica* 80(4), 1337–85.
- Bronnenberg, B. J. and J.-P. Dubé (2017). The formation of consumer brand preferences. *Annual Review of Economics* 9, 353–82.
- Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In James Heckman and Barton Singer, (Eds.), *Longitudinal Analysis of Labor Market Data*, 3–38. *Econometric Society Monographs*. Cambridge University Press.
- Chernozhukov, V., I. Fernandez-Val, J. Hahn and W. Newey (2013). Average and quantile effects in non-separable panel models. *Econometrica* 81(2), 535–80.
- Chesher, A. and J. Santos Silva (2002). Taste variation in discrete choice models. *Review of Economic Studies* 69(1), 147–68.
- Ching, A. T. (2010). Consumer learning and heterogeneity: dynamics of demand for prescription drugs after patent expiration. *International Journal of Industrial Organization* 28, 619–38.
- Cox, D. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B* 20(2), 215–42.
- Davezies, L., X. D’Haultfoeuille and L. Laage (2021). Identification and estimation of average marginal effects in fixed effect logit models. Working paper, arXiv Preprint: 2105.00879.
- De Groote, O. and F. Verboven (2019). Subsidies and time discounting in new technology adoption: evidence from solar photovoltaic systems. *American Economic Review* 109(6), 2137–72.
- Dobronyi, C., J. Gu and K. Kim (2021). Identification of dynamic panel logit models with fixed effects. Working paper, arXiv Preprint: 2104.04590.
- Erdem, T., S. Imai and M. Keane (2003). Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics* 1(1), 5–64.
- Esteban, S. and M. Shum (2007). Durable-goods oligopoly with secondary markets: the case of automobiles. *RAND Journal of Economics* 38(2), 332–54.
- Gandhi, A. and A. Nevo (2021). Empirical models of demand and supply in differentiated products industries. In *Handbook of Industrial Organization Volume 4*, 63–139. Elsevier.
- Goettler, R. L. and B. R. Gordon (2011). Does amd spur intel to innovate more? *Journal of Political Economy* 119(6), 1141–200.
- Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy* 120(6), 1173–219.
- Griffith, R., L. Nesheim and M. O’Connell (2018). Income effects and the welfare consequences of tax in differentiated product oligopoly. *Quantitative Economics* 9(1), 305–41.
- Guadagni, P. and J. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3), 203–38.
- Hahn, J. (2001). The information bound of a dynamic panel logit model with fixed effects. *Econometric Theory* 17(5), 913–32.
- Heckman, J. (1981a). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Daniel McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, 179–95. MIT Press: Cambridge, MA.
- Heckman, J. (1981b). Statistical models for discrete panel data. In *Structural Analysis of Discrete Data with Econometric Applications*. Charles Manski and Daniel McFadden, (Eds.), 114–78. MIT Press: Cambridge, MA.
- Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica* 74(6), 1637–73.
- Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review* 103(7), 2722–51.

- Hitsch, G. J., A. Hortacsu and X. Lin (2019). Prices and promotions in us retail markets: evidence from big data. Technical report, National Bureau of Economic Research, NBER Working paper, 26306.
- Honoré, B. and A. De Paula (2021). Identification in simple binary outcome panel data models. *Econometrica* 24(2), C78–C93.
- Honoré, B. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* 68(4), 839–74.
- Honoré, B. and E. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74(3), 611–29.
- Honoré, B. and M. Weidner (2020). Dynamic panel logit models with fixed effects. Working paper, arXiv preprint: 2005.05942.
- Honoré, B. E., C. Muris and M. Weidner (2021). Dynamic ordered panel logit models. Working paper, arXiv preprint: 2107.03253.
- Jones, J. M. and J. T. Landwehr (1988). Removing heterogeneity bias from logit model estimation. *Marketing Science* 7(1), 41–59.
- Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77(1), 135–75.
- Keane, M. (2015). Panel data discrete choice models of consumer demand. In *The Oxford Handbook of Panel Data*, Badi H. (Eds.), 548–82. Oxford University Press. Oxford, UK.
- Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics* 15(3), 310–27.
- Kitazawa, Y. (2013). Exploration of dynamic fixed effects logit models from a traditional angle. Technical report, Kyushu Sangyo University, Working paper, 60.
- Kitazawa, Y. (2022). Transformations and moment conditions for dynamic fixed effects logit models. *Journal of Econometrics* 229(2), 350–62.
- Kong, X., J.-P. Dubé and Ø. Daljord (2022). Nonparametric estimation of habitual brand loyalty, Working paper, Available at SSRN.doi.org/10.2139/ssrn.4070747.
- Liu, Y. and S. Balachander (2014). How long has it been since the last deal? Consumer promotion timing expectations and promotional response. *Quantitative Marketing and Economics* 12(1), 85–126.
- Mysliwski, M., F. M. Sanches, D. S. Junior and S. Srisuma (2020). The welfare effects of promotional fees. CeMMAP Working Paper, CWP35/20.10.1920/wp.cem.2020.3520
- Nevo, A. (2011). Empirical models of consumer behavior. *Annual Review of Economics* 3(1), 51–75.
- Nielsen, NielsenIQ (2022). Retail scanner data from NielsenIQ at the Kilts center for marketing, University of Chicago Booth School of Business. <https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen#simple2>.
- Osborne, M. (2011). Consumer learning, switching costs, and heterogeneity: a structural examination. *Quantitative Marketing and Economics* 9(1), 25–70.
- Pakel, C. and M. Weidner (2021). Bounds on average effects in discrete choice panel data models. Technical report, Working paper.
- Pavlidis, P. and P. B. Ellickson (2017). Implications of parent brand inertia for multiproduct pricing. *Quantitative Marketing and Economics* 15(4), 369–407.
- Pesendorfer, M. (2002). Retail sales: a study of pricing behavior in supermarkets. *Journal of Business* 75(1), 33–66.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research. Series: Studies in Mathematical Psychology.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 4, 321–33.

- Roy, R., P. Chintagunta and S. Haldar (1996). A framework for investigating habits, 'the hand of the past', and heterogeneity in dynamic brand choice. *Marketing Science* 15(3), 280–99.
- Ryan, S. P. and C. Tucker (2012). Heterogeneity and the dynamics of technology adoption. *Quantitative Marketing and Economics* 10(1), 63–109.
- Seetharaman, P., A. Ainslie and P. Chintagunta (1999). Investigating household state dependence effects across categories. *Journal of Marketing Research* 36(4), 488–500.
- Shum, M. (2004). Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market. *Journal of Economics & Management Strategy* 13(2), 241–72.
- Simonov, A., J.-P. Dubé, G. Hitsch and P. Rossi (2020). State-dependent demand estimation with initial conditions correction. *Journal of Marketing Research* 57(5), 789–809.
- Torgovitsky, A. (2019). Nonparametric inference on state dependence in unemployment. *Econometrica* 87(5), 1475–1505.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Supplement

*Co-editor Jaap Abbring handled this manuscript.*