Non-parametric Estimation of Habitual Brand Loyalty

Xinyao Kong, Chicago Booth *

Jean-Pierre Dubé, Chicago Booth and NBER

Øystein Daljord, Chicago Booth

March 8, 2022

Abstract

US companies spend billions on point-of-sale promotions that have prompted decades of research devoted to their rationalization as a response to dynamic habitual brand loyalty (HBL) in consumer behaior. We propose a nonparametric test for HBL using a "dynamic potential outcomes" model that resovles the classic identification challenge of decoupling state dependence and unobserved heterogeneity. We then propose a semi-parametric test for forward-looking behavior to assess whether consumers rationally plan their HBL. Case studies of several large consumer packaged goods categories reveal non-trivial extent of HBL in consumers' brand choices. We also find semi-parametric evidence for forward-looking consumer behavior and strongly reject the usual parametric static discrete-choice model in favor of the dynamic discrete-choice model with a freely-varying discount factor. The long-run price elasticities from a dynamic discrete-choice model are found to be considerably larger in magnitude than those from a model with myopic choices, with substitution patterns often more than 1000% higher in magnitude.

*Authors contributed equally and names are listed in reverse alphabetical order. E-mail: xinyao.kong@chicagobooth.edu, jean-pierre.dube@chicagobooth.edu. We thank IRI for providing the IRI Academic Database. The opinions expressed in this paper do not reflect those of the data providers. We are extremely grateful to Alex Torgovitsky for comments feedback. We are also grateful to the comments and suggestion of Avner Shlain and seminar participants at the Booth School of Business, the University of Illinois at Urbana-Champaign, the University of Michigan, the European Virtual Quantitative Marketing Seminar, and the Virtual Quantitative Marketing Seminar. Dubé acknowledges research support from the Kilts Center for Marketing and the Charles E. Merrill faculty research fund.

1 Introduction

US companies spent $240 billion on advertising in 2020,[1]. Companies allocate about 75% of this budget to sales promotions geared largely towards stealing market share from competitors,Sales promotions typically consist of temporary discounts, coupons, in-store displays, loyalty programs and newspaper feature ads geared towards influencing the consumer at the point-of-sale. up significantly from the mid-20th century when promotions represented less than half of advertising spending (Jones 1990). Tracking the growth in share-stealing promotions, a long literature in quantitative marketing has studied the positive question of whether consumers form brand habits and exhibit brand loyalty, mechansims that could rationalize promotional investments. Indeed, since the advent of transaction-level shopping panels, the literature has documented high rates of brand repeat-purchase within-consumer despite heterogeneous brand choices between consumers (e.g., Brown, 1953; Cunningham, 1956; Kuehn, 1958). Of interest is whether the high repeat-purchase rates reflect *habitual brand loyalty* (HBL): the treatment effect of past choices on current choices (e.g., see Bass et al., 1984 for a survey). HBL is typically believed to arise from psychological *switching costs* (e.g., Klemperer, 1995; Farrell and Klemperer, 2007; Dubé et al., 2010) that can generate substantial barriers to entry into consumer goods industries (Bain, 1956). HBL may also account for some of the $235 billion in estimated annual U.S. intangible marketing capital (Corrado et al., 2005). Furthermore, HBL is one of the leading justifications for potential ROI from the $244.7 billion spent on sales promotion for consumer goods at the point-of-sale.[2] However, the extant empirical literature has generated mixed results for HBL with inconclusive evidence from under-powered, non-parametric approaches (e.g., Frank, 1962; Massy, 1966; Bass et al., 1984) and parametric evidence based on random utility models that rely on ad hoc parametric assumptions (e.g., Keane, 1997; Seetharaman et al., 1999; Dubé et al., 2010).

We apply a novel approach proposed by Torgovitsky (2019) that re-casts HBL within a non-parametric Dynamic Potential Outcomes (DPO) framework. The model relates observed choice outcomes to potential outcomes that would have been observed had a past choice been exogenously

[1] Accessed from Statista on 3-6-2022 at https://www.statista.com/statistics/429036/advertising-expenditure-in-north-america/: :text=The%20United%20States%20is%20the,billion%20U.S.%20dollars%20in%202020.

[2] These expenditures include temporary discounts and merchandizing and displays to promote the sale of a product in a store. Numbers accessed at https://www.statista.com/statistics/987009/marketing-spending-us-category/ on March 3, 2022.

assigned. HBL – the causal effect of past choices on current choices – exists if potential outcomes would be different under different hypothetical past choices. The DPO notation also clarifies the separation between brand loyalty and unobserved heterogeneity (e.g., Heckman, 1981; Heckman and Singer, 1984; Honore and Kyriazidou, 2000; Honoré and Tamer, 2006; Torgovitsky, 2019; Pakes et al., 2021) and is not vulnerable to the "initial conditions" problem (e.g., Heckman, 1981; Simonov et al., 2020). The DPO also nests the canonical random utility model (RUM) as well as its dynamic discrete-choice model (DDC) analog. As a result, the DPO approach enables us to set-identify HBL using behavioral restrictions implied by random utility theory without literally estimating a parametric RUM model.

We then extend this positive analysis to test for forward-looking behavior in the presence of brand loyalty to assess whether consumers consciously plan their future brand loyalty. We show that the standard DDC with HBL has *built-in* exclusion restrictions that satisfy the conditions in Abbring and Daljord (2020) to identify the discount factor and the utility function jointly. We then propose a semi-parametric minimum distance estimator for the discount factor. We also explore a parametric dynamic-discrete-choice model of demand with HBL that includes the discount factor as a freely-varying parameter. Our maximum likelihood estimator is formulated as a MPEC problem ((Su and Judd, 2012; Dubé et al., 2012) that imposes Bellman's equation as constraints.

We conduct several empirical case studies of CPG markets using the IRI Academic Data Set, focusing on categories with well-established brands to minimize concerns about short-term sources of state-dependence such as learning. The data combine household-level shopping panel data and the point-of-sale marketing conditions for the stores in which the panelists shop. Using the DPO framework, we find evidence for HBL in brand choices for several of the CPG categories studied. The lower bounds on HBL for the leading brands in these categories are often found to be well above zero. In particular, for several categories, we find that as much as 25% of the current purchases of the top-selling brand is due to HBL. We also find several cases where over 25% of the choice-persistence for the top-selling brand (i.e., repeat purchases) is due to HBL. Interestingly, our upper bounds are also quite sharp with most categories exhibiting less than 50% HBL, suggesting that this form of brand loyalty may not be as pervasive as would be implied in parametric choice models.

We also find strong evidence of consumer patience, with the mean discount factor in many categories close to the value implied by the real rate of interest. These results confirm that the expected future value helps fit the moments implied by our exclusion restriction, suggesting that consumers consciously plan their future. Nested and non-nested tests reject the standard, myopic random-utility model of choice in favor of the dynamic model.

Our findings have important normative implications for pricing and marketing. In particular, we find considerably larger equilibrium elasticities using a DDC with brand loyalty and a freely-varying discount factor parameter versus a static RUM with brand loyalty and myopic behavior. The cross-price elasticities of demand are often found to be over 1000% higher in the DDC, implying considerably more competition and susbstitution once forward-looking behavior is accounted for. These results echo the findings of Hendel and Nevo (2006)) who also find large differences between the elasticities from the myopic and forward-looking choice model; although we find a systematic downward bias in price-sensitivity when the myopic choice model with HBL is used. In addition, we test the assumption of myopia and estimate a discount factor, as opposed imposing forward-looking subject to a deterministically-known discount factor.

Our work contributes to the broader economic literature estimating state-dependent versions of the static RUM to test for and measure switching costs in the healthcare industry (e.g., Handel (2013),Polyakova (2016). The DPO formulation allows for a more agnostic test of the extent to which consumer choices are driven by switching costs. To the best of our knowledge, this literature assumes myopic behavior. We see high potential value to the application of our semi-parametric estimator of the discount factor to test for rational forward-looking consumer behavior in the choice of healthcare plans and whether consumers are conscious of the long-term implications of their choices. More generally, the test for forward-looking behavior also contributes to the small, but growing literature studying the underlying mechanisms governing HBL (e.g., Farrell and Klemperer, 2007; Dubé et al., 2010; Gordon and Sun, 2015).

Our approach is most closely related to the recent study of switching costs by Pakes et al. (2021), who propose a moment inequalities estimator that generates semi-parametric bounds on a linear HBL parameter. In contrast, our DPO evidence for HBL is fully non-parametric, making it robust to a broad class of parametric choice-specific index function and to the underlying consumer choice behavior (e.g., it nests dynamic discrete-choice as well as boundedly-rational forms of behavior). It is unclear whether the approach in Pakes et al. (2021) could be adapted to the case of dynamic discrete-choice, with bounds on the consumer discount factor.

Our work is also related to an older literature testing for HBL using statistical tests applied only to within-panelist variation, such as the non-parametric binomial runs test and the parametric t-test for higher-order, Markovian choice outcomes (e.g., Frank, 1962; Massy, 1966; Bass et al., 1984). The DPO approach is more agnostic as it only imposes restrictions on the potential outcomes path rather than imposing restrictions on the actual probability distribution governing the observed choice outcomes. For instance, a binomial runs test uses a stationary binomial process as the null model. As a result, the DPO approach is also more robust to the potential for latent marketing variables, such as prices and advertising, that may also affect choices. In addition, the DPO formulation has more statistical power in typically short-duration shopping panels because it allows cross-sectional pooling of households in addition to using the within-household variation. Finally, the DPO model bounds the extent of HBL in the population rather than just testing for incidence of higher-order choice behavior.

The rest of the paper is organized as follows. Section 2 defines the canonical discrete-choice formulation of brand purchase behavior. Section 3 discusses the DPO model and the empirical bounds on HBL, along with a set of identifying assumptions to sharpen the bounds. Section 4 derives moments from the DDC to identify and estimate the discount factor. Section 5 describes our empirical case studies and results, with policy implications discussed in section 6. We conclude in section 7.

2 Habitual Brand Loyalty and the Canonical Random Utility Model

We start by presenting the canonical, discrete-choice random utility model with unit demand (RUM) and habitual brand loyalty (HBL). Each time period $t = 1,...,T$, a consumer makes a discrete purchase decision $d_t \in \{0,1,\ldots,J\} \equiv \mathscr{D}$ where $\mathscr{D}$ is a choice set containing a no-purchase alternative, $j = 0$, and $j = 1,...,J$ brands.

A consumer in state $s \in \mathscr{S}$ faces choice-specific utility is $u_j(s_t) + \varepsilon_{jt}$, $\forall j \in \mathscr{D}$ where we make the conventional normalization, $u_{J+1}(s_t) = 0$ and where $\varepsilon_t \equiv (\varepsilon_{0t},...,\varepsilon_{Jt})' \sim$ i.i.d. $F_\varepsilon(\varepsilon)$ is a $(Jx1)$

vector of random utility disturbances. We assume $F_{\varepsilon}(\varepsilon)$ is the Type-I Extreme value distribution.

The state vector $s_t = (\mathbf{p}_t, \ell_t)$ contains the vector of current prices, $\mathbf{p}_t \equiv (p_{1t}, ..., p_{Jt})$, and the consumer's current loyalty state, $\ell_t \in \{1, \ldots, J\} \equiv \mathscr{D}_{-0}$, reflecting her brand choice history. The consumer cannot be loyal to non-purchase, $j = 0$. We assume the consumer's loyalty state evolves according to a controlled Markov process with transition probabilities $\Pr(\ell_{t+1} = k | \ell_t, d_t), \forall k \in \mathscr{D}_{-0}$ where $d_t \in \mathscr{D}$ is the period-t purchase decision. We assume that $\mathbf{p}_t$ evolves according to an exogenous Markov process with discrete support, $\mathscr{P}$. The state space $\mathscr{S} = \mathscr{P} \times \mathscr{D}_{-0}$ is, therefore, discrete with L support points. We denote the consumer's beliefs about the evolution of the state as the controlled Markov transition distribution $\mathbf{F}_{d_t}(s_t) = [\Pr(s_{t+1} = s_1 | s_t, d_t), \ldots, \Pr(s_{t+1} = s_L | s_t, d_t)]$.

The ex-ante value function associated with the consumer's purchase decision problem is

$$v(s_t) = \mathbb{E}_{\varepsilon} \left[\max_{j \in \mathscr{D}} \left\{ v_j(s) + \varepsilon_j \right\} \right] \tag{2.1}$$

where the choice-specific value functions are defined as the solutions to

$$v_j(s) \quad = \quad u_j(s_t) + \beta \mathbf{F}_j(s) \ln \left\{ \sum_{k \in \mathscr{D}} \exp(\mathbf{v}_k) \right\}, \ \forall j \in \mathscr{D} \tag{2.2}$$

where $\beta \in (0,1)$ is a discount factor, and $\mathbf{v}_j = [v_j(s_1), \ldots, v_j(s_L)]'$ is an $L \times 1$ vector of values from choosing brand j in each of the L states. For notational convenience, we can re-write the choice-specific value functions in (2.2) in the following matrix form:

$$v_j(s) \quad = \quad u_j(s_t) + \beta \mathbf{F}_j(s) (\mathbf{m} + \mathbf{v}_0),$$

where $\mathbf{v}_0 = [v_0(s_1), \ldots, v_0(s_L)]'$ and $\mathbf{m} = [-\ln(\sigma_0(s_1)), \ldots, -\ln(\sigma_0(s_L))]'$ are $L \times 1$ vectors. This is the classic dynamic-discrete-choice model (DDC) (e.g., Rust (1987)).

The consumer's utility-maximizing choice probabilities in state s are:

$$\sigma_j(s) = \frac{\exp(v_j(s))}{\sum_{k \in \mathscr{D}} \exp(v_k(s))}, \ \ \forall j \in \mathscr{D}. \tag{2.3}$$

The extant literature typically tests for HBL with variations of the following hypothesis:

$$\mathbb{H}_0 : HBL_j \equiv v_j(\mathbf{p}, \ell = j) - v_j(\mathbf{p}, \ell \neq j) > 0, \ \forall j \in \mathscr{D}_{-0}.$$

The alternative hypotheses consist of *variety-seeking* ($HBL < 0$) and *no state-dependence* ($HBL = 0$).

3 The Dynamic Potential Outcomes Model

We use the non-parametric, Dynamic Potential Outcomes model (DPO) (Torgovitsky, 2019) to reformulate HBL as the causal effect of past choices on current choices. Our discussion and notation follow Torgovitsky (2019) with one important distinction: we extend the DPO to the *multinomial* brand-choice context. The DPO nests the RUM with HBL from Section 2 without requiring any parametric assumptions on preferences or the discount function. Furthermore, the DPO model set identifies HBL in brand purchases using mild, economically-motivated identifying assumptions, such as stationarity and monotonicity, that are implied by the RUM.

3.1 Model

A consumer i makes observed brand-choice decisions $Y_{it} \in \mathscr{J}$ in time periods $t = 1, ..., T$. We exclude the no-purchase alternative from this discussion. To capture the state-dependent effect of brand loyalty, we assume the consumer receives the *treatment* $Y_{i(t-1)}$ in period t so that brand loyalty is defined as the causal effect of last period's brand-choice on the current period's brand choice outcome. Each period t, there exists a set of unobserved potential outcomes, $U_{it}(1), ..., U_{it}(J) \in \mathscr{J}$, corresponding to the brand choices that would have occurred under each of the loyalty states $Y_{i(t-1)} \in \mathscr{J}$.[3]

The observed brand-choice outcomes, $\mathbf{Y}_i \equiv (Y_{i0}, Y_{i1}, ..., Y_{iT}) \in \mathscr{Y} \equiv \{1, ..., J\}^T$ are related to potential outcomes $\mathbf{U}_i(1) \equiv (U_{i1}(1), ..., U_{iT}(1))$, ..., $\mathbf{U}_i(J) \equiv (U_{i1}(J), ..., U_{iT}(J))$ through the DPO

[3] This first-order structure of loyalty does not impose a temporal structure on the potential outcomes, such as a first-order Markov chain. It would also be straightforward to redefine the potential outcomes over sequences of past choices to allow for richer forms of loyalty.

model

$$Y_{it} = \sum_{j=1}^{J} \mathbb{1}[Y_{i(t-1)} = j] U_{it}(y), \; \forall t \geq 1 \tag{3.1}$$

where the initial condition Y_{i0} is observed but not modeled, avoiding the *initial conditions* problem Heckman (1981); Honoré and Tamer (2006).

In contrast with past research that defines HBL through the observed outcomes, $\mathbf{Y}_i$, the DPO defines HBL through the unobserved potential outcomes $\mathbf{U}_i$. The DPO allows for state dependence through the probability that different loyalty states cause different potential outcomes. With no further restrictions, the DPO exhibits HBL when we observe positive state-dependence for a brand j: $U_{it}(j) = j,\, U_{it}(k) \neq j,\; j \neq k \in \mathscr{J}$. The DPO exhibits variety-seeking when we observe negative state-dependence for a brand j: $U_{it}(j) \neq j,\, U_{it}(k) = j,\; j \neq k \in \mathscr{J}$. The loyalty state has no causal effect in the DPO when we observe the same potential outcomes for a brand j in different loyalty states: $U_{it}(j) = U_{it}(k),\; j \neq k \in \mathscr{J}$, such as when brand choice reflects (persistent) tastes.

A classic challenge in the literature on state-dependence in choices is the separate identification of state-dependence and heterogeneity Frank (1962); Heckman (1981); Keane (1997); Dubé et al. (2010). The DPO also captures unobserved heterogeneity by allowing variation in the potential outcomes $(U_{it}(1), ..., U_{it}(J))$ in the population. To illustrate, consider a two-period, binary choice example in which brand 1 is repeatedly purchased: $\mathbf{Y}_i = (Y_{i0}, Y_{i1}) = (1,1)$, where $Y_{it} \in \{0,1\}$ for $t = 0,1$. As illustrated in Figure 1, any realization of potential outcomes with $U_{i1}(1) = 1$ can generate the choice sequence $\mathbf{Y}_i$. Therefore, the DPO could explain $\mathbf{Y}_i$ as preferences (i.e., $U_{i1}(0) = U_{i1}(1) = 1$) or as HBL (i.e., $U_{i1}(0) = 0$, and $U_{i1}(1) = 1$).

3.2 The DPO Model Nests the Canonical Random Utility Model

The DPO nests the static and dynamic versions of the RUM from section 2. In the canonical RUM, the potential outcomes in any given state s are the choices generated by utility maximization

$$U_{it}(y) = \arg\max_{k \in \mathscr{J}} \{v_k(y) + \varepsilon_{kt}\}, \; \forall y \in \mathscr{J} \tag{3.2}$$

$$
\begin{array}{c|cccc}
Y_{i1} & 1 & 1 & 1 & 1 \\
\hline
U_{i1}(0) & 1 & 0 & 1 & 0 \\
U_{i1}(1) & \boxed{1} & \underbrace{\boxed{1}}_{\text{state dep}} & 0 & 0
\end{array}
$$

The columns in black are the potential outcomes that could generate observed choices $\mathbf{Y}_i = (1,1)$ through the recursive relationship (3.1). The columns in gray are those potential outcomes that could not generate observed choices. The squared cells are potential outcomes in the observed loyalty state, while the rest are unobserved potential outcomes.

Figure 1: Decomposing observed choice persistence

where the treatment is $Y_{t-1} \in \mathscr{J}$. As before, we can relate the observed outcomes $\mathbf{Y}_i$ to the potential outcomes as follows:

$$
Y_{it} = \sum_{y \in \mathscr{J}} \mathbb{1}[Y_{i(t-1)} = y] U_{it}(y). \tag{3.3}
$$

Recall that the RUM in section 2 includes a no-purchase alternative, $j = 0$, to which consumers do not form any loyalty. For the empirical DPO analysis, we condition on purchase (i.e., exclude non-purchase) to reduce the computational burden. However, it is straightforward to reformulate 3.3 to accommodate no-purchase in our DDC analysis as follows:

$$
\begin{aligned}
U_{it}(y) &= \arg\max_{k \in \mathscr{J}} \{v_k(y) + \varepsilon_{kt}\},\ \forall y \in \mathscr{D} \\
Y_{it} &= \sum_{y \in \mathscr{J}} \mathbb{1}[Y_{i(t-1)} = y] U_{it}(y) + \mathbb{1}[Y_{i(t-1)} = 0] U_{it}(\ell_{t-1}) \\
\ell_t &= \mathbb{1}[Y_{i(t-1)} = 0]\ell_{t-1} + \mathbb{1}[Y_{i(t-1)} \neq 0] Y_{i(t-1)}
\end{aligned} \tag{3.4}
$$

where y is a counterfactual loyalty state, and ℓ_t is the realized loyalty state.

While the DPO model is agnostic about the parametric form of the potential outcomes function $U_{it}(\cdot)$, the RUM model imposes several parametric restrictions including utility maximization behavior, the additive separability of the random utility disturbances, ε, and a parametric distributional assumption for the utility disturbances, $F_\varepsilon(\varepsilon)$. In the case of the DDC, the analysis further assumes geometric discounting and knowledge of the consumers' beliefs about the state transition process.

3.3 Identification of HBL in the DPO model

The empirical goal consists of testing for a causal effect of past brand choice on current decision-making. We therefore define our target parameter as follows:

$$\mathbf{HBL}_t(P) = (HBL_{1t}(P), ..., HBL_{Jt}(P)), \tag{3.5}$$

where

$$HBL_{jt}(P) = \mathbb{P}_P[\exists k \neq j \in \mathscr{J} \ \ s.t. \, U_{it}(j) = j, U_{it}(k) \neq j] \tag{3.6}$$

where P denotes the probability mass function over the possible potential outcome paths, $\mathbf{U}_i \equiv (Y_{i0}, \mathbf{U}_i(1), ..., \mathbf{U}_i(J))$, with support $\mathscr{U} \equiv \{1, .., J\}^{JT+1}$. Formally, HBL_{jt} measures the population probability on date t that there exists at least one brand $k \neq j$ such that a consumer who would choose brand j if loyal to it would choose something else if she had instead been loyal to k.

Measuring HBL under the DPO model (3.1) differs conceptually from the more familiar RUM framework, where HBL consists of a preference parameter. Our model primitive now consists of the probability mass function P, where $P(\mu)$ represents the probability of observing potential outcome path μ in the consumer population. Let $\mathscr{P}$ denote the set of all probability mass functions P where

$$P : \mathscr{U} \to [0, 1] \quad \text{s.t.} \sum_{\mu \in \mathscr{U}} P(\mu) = 1. \tag{3.7}$$

The parameter space, $\mathscr{P}^\dagger$, consists of the subset of $\mathscr{P}$ that satisfies the researcher's prior assumptions:

$$\mathscr{P}^\dagger = \{P \in \mathscr{P} : \rho(P) \geq 0\}, \tag{3.8}$$

where $\rho : \mathscr{P} \to \mathbb{R}^{d_\rho}$ is a function representing assumptions on P, and the inequality is interpreted component-wise. $\rho(P)$ denotes addtional prior restrictions, discussed below in more detail, that we will impose restrictions to improve identification. In the empirical application of the DPO, we will also need to address the fact that the dimension of $\mathscr{U}$ increases exponentially in J and T.

We can now define the *identified set*, $\mathscr{P}^\star$, as the observationally equivalent subset of the parameter space, $\mathscr{P}^\dagger$, for which the DPO can rationalize the observed brand choices. Let $\mathbb{P}[\mathbf{Y}_i = y]$ denote the observed probability of $\mathbf{Y}_i$. Then the observational equivalence requires that, for every $y \equiv (y_0, y_1, .., y_T) \in \mathscr{Y}$,

$$\begin{aligned} \mathbb{P}[\mathbf{Y}_i = y] &= \mathbb{P}_P[Y_{i0} = y_0, U_{it}(y_{t-1}) = y_t \text{ all } t \geq 1] \\ &= \sum_{\mu \in \mathscr{U}_{oeq}(y)} P(\mu), \end{aligned} \tag{3.9}$$

where $\mathscr{U}_{oeq}$ is the set of all potential outcome sequences that can generate the observed choices $y = (y_0, y_1, .., y_T)$ through $\mu_0 = y_0$ and $\mu_t(y_{t-1}) = y_t$ for all $t \geq 1$.

We now turn to the testing of HBL. Suppose we define a target parameter $\theta : \mathscr{P} \to \mathbb{R}^{d_\theta}$ as a low dimensional function of P. The researcher is interested in the identified set for θ, denoted by $\Theta^\star \equiv \{\theta(P) : P \in \mathscr{P}^\star\}$. An example of the target parameter θ is $HBL_{jt}(P)$ defined above. Recall that we only observe the potential outcome in the observed state, $U_{it}(Y_{i(t-1)})$, so that $\{U_{it}(y)\}_{y \in \mathscr{J}}$ is unobserved. This is the heart of the identification problem and the following proposition clarifies that the HBL_{jt} parameters (3.6) are not point-identified but are set-identified. The proposition is an extension of Torgovitsky (2019)'s Proposition 1 to the multinomial context.

Proposition 1. *Suppose that $\mathscr{P}^\dagger = \mathscr{P}$. If $\theta = HBL_{jt}(P)$, then*

$$\begin{aligned} \Theta^\star &= \big[0, \mathbb{P}[Y_{i(t-1)} = j, Y_{it} = j] + \mathbb{P}[Y_{i(t-1)} \neq j]\big] \quad \textit{when } |\mathscr{J}| \geq 3, \\ \Theta^\star &= \big[0, \mathbb{P}[Y_{i(t-1)} = j, Y_{it} = j] + \mathbb{P}[Y_{i(t-1)} \neq j, Y_{it} \neq j]\big] \quad \textit{when } |\mathscr{J}| = 2. \end{aligned} \tag{3.10}$$

See Appendix $A.7$ for the proof. Proposition 1 confirms that in the absence of any prior restrictions (i.e., $\mathscr{P}^\dagger = \mathscr{P}$), the data alone cannot reject pure heterogeneity even though it can bound the extent of state dependence from above. The upper bounds also show us how variation in choices identify the bounds. If consumers never switched brands, then the upper bound for HBL would be 100% and there would be no empirical content to the assumption free bounds.

Clearly we will need to impose more restrictions to sharpen the bounds in Proposition 1, as we discuss below in section 3.3.1. First, we present a simple pedagocial example to illustrate the identification result.

3.3.1 A Pedagogical Example

To illustrate the identification of **HBL**, consider the following two-consumer, three-period example. Assume there are two consumers, $i = 1,2$, in the population and three time periods $t = 0,1,2$. Suppose we observe the choice sequences $\mathbf{Y}_1 = (0,0,1)$ and $\mathbf{Y}_2 = (0,1,1)$ for consumers 1 and 2, respectively. The corresponding observed choice probabilities are

$$\mathbb{P}[\mathbf{Y}_i = (0,0,1)] = 1/2$$
$$\mathbb{P}[\mathbf{Y}_i = (0,1,1)] = 1/2$$

Let $\mu = (\mu_0, (\mu_1(0), \mu_1(1)), (\mu_2(0), \mu_2(1)))$ denote a potential outcomes path. The set $\mathscr{U} = \{0,1\}^{2*2+1}$ contains all 32 possible potential outcomes paths. We denote the set of observationally equivalent potential outcome paths that rationalize the data as $(0,(0,\cdot),(1,\cdot))$ and $(0,(0,\cdot),(1,\cdot))$, where

$$P[\mu \in (0,(0,\cdot),(1,\cdot))] = \mathbb{P}[\mathbf{Y}_i = (0,0,1)] = \frac{1}{2} \tag{3.11}$$
$$P[\mu \in (0,(1,\cdot),(\cdot,1))] = \mathbb{P}[\mathbf{Y}_i = (0,1,1)] = \frac{1}{2} \tag{3.12}$$

Only 8 of the 32 potential outcome paths can be consistent with the data. However, we cannot fully rule out pure heterogeneity since, for example, $\mu = (0,(0,1),(1,1))$ is consistent with $\mathbf{Y}_1$ and its potential outcomes in period $t = 2$ implies pure heterogeneity.

Therefore, without additional restrictions, the identified set corresponding to our data is

$$\begin{aligned}\mathscr{P}^{\star}_{oeq} = \{P: \quad & \sum_{\mu \in \mathscr{U}} P(\mu) = 1, \quad P(\mu) \in [0,1] \\ & P[\mu \in (0,(0,\cdot),(1,\cdot))] = \mathbb{P}[\mathbf{Y}_i = (0,0,1)] = \frac{1}{2} \\ & P[\mu \in (0,(1,\cdot),(\cdot,1))] = \mathbb{P}[\mathbf{Y}_i = (0,1,1)] = \frac{1}{2}\}\end{aligned}$$

where the first row is required for P to be a probability mass function on $\mathscr{U}$, and the rest are required for observational equivalence.

We can now define the target parameters $\mathbf{HBL}(P) = (HBL_1(P), HBL_2(P))$. For period 1, we have

$$\begin{aligned}HBL_1(P) &= \mathbb{P}_P[\mu \in (\cdot,(0,1),(\cdot,\cdot))] \\ &= P[\mu = (0,(0,1),(1,0))] + P[\mu = (0,(0,1),(1,1))] \end{aligned} \tag{3.13}$$

where the second equality follows from the fact that only two of the 8 potential outomce paths consistent with $\mathbf{Y}_1$ and $\mathbf{Y}_2$ exhibit HBL in period 1. With no additional restrictions, we can bound $HBL_1(P)$ within the closed interval $[0,1/2]$. To see this result, the two potential outcome paths in HBL_1 are consistent with $\mathbf{Y}_1$ and so are bounded by (3.11),

$$0 \leq P[\mu = (0,(0,1),(1,0))] + P[\mu = (0,(0,1),(1,1))] \leq \mathbb{P}[\mathbf{Y}_i = (0,0,1)] = 1/2$$

Similarly HBL_2 is also bounded within $[0,1/2]$.[4]

We've shown that the identified sets for habitual brand loyalty in the data are $H\hat{B}L_1 = H\hat{B}L_2 = [0,1/2]$. The upper bounds of HBL_t depend on the extent of serial dependence in choices: consumer 1's repeated choice of 0 in $t = 0,1$ bounds HBL_1 from above and consumer 2's repeated choice of 1 in $t = 1,2$ bounds HBL_2 from above. However, the data alone fail to reject "no HBL" in the population in any period.

[4] $HBL_2 = P[\mu = (0,(1,0),(0,1))] + P[\mu = (0,(1,1),(0,1))]$; $0 \leq HBL_2 \leq \mathbb{P}[\mathbf{Y}_i = (0,1,1)] = 1/2$.

3.4 Identifying Assumptions

We now follow Torgovitsky (2019) and propose additional restrictions to sharpen the bounds on the identified sets for HBL. As discussed earlier, the RUM imposes several parametric restrictions that ensure the point identification of the model parameters. However, these parametric assumptions may be unnecessary to test for HBL. We instead introduce three behaviorally-motivated identifying assumptions implied by the RUM, including stationarity and monotonicity assumptions. The assumptions can be added to the DPO model via the ρ function in (3.8) to sharpen bounds on the HBL parameters.

3.4.1 Stationarity

We can use a stationarity assumption to pool the population distribution of potential outcomes from different time periods. A simple form of stationarity is as follows:

Assumption ST: $\forall \mu \equiv (\mu(1), ..., \mu(J)) \in \{1, ..., J\}^J, \forall t, t' \geq 1,$

$$\mathbb{P}_P[U_{it}(1) = \mu(1), ..., U_{it}(J) = \mu(J)] = \mathbb{P}_P[U_{it'}(1) = \mu(1), ..., U_{it'}(J) = \mu(J)]$$

This assumption ensures the time-invariance of the joint distribution of potential outcomes $(U_{it}(1), ..., U_{it}(J))$ in the population.

Unlike the early *stochastic brand choice* literature (e.g., see the survey in Bass et al. (1984)) that imposed stationarity on the within-consumer *observed* outcomes Y_{it}, the DPO model imposes stationarity on the cross-sectional distribution of *potential* outcomes U_{it}. When stationarity is imposed on potential outcomes, a sufficient condition for stationarity in observed outcomes is that there is no state dependence in the data generating process. Therefore, assuming Stationarity makes it possible to bound HBL away from 0 if we observe non-stationary choices.

Although the DPO model itself is agnostic about the mechanism of state dependence in brand choices, imposing Stationarity assumes away mechanisms that generate non-stationary treatment effect of past choices on current choices as the sole source of such state dependence. As an example, Bayesian learning cannot be the only reason that brand choices exhibit state dependence under Assumption ST, because its treatment effect of the past choice on the current choice depends on the number of signals sampled and hence is non-stationary.

The following proposition establishes that (ST) is implied by the standard RUMs with time-invariant coefficients along with stationary covariates and utility shocks.

Proposition 2. *Suppose that every $P \in \mathscr{P}^{\dagger}$ is consistent with $U_{it}(y)$ being generated through (3.3) by an RUM model. If the distributions of prices, $\mathbf{p}_t$, and random utility shocks, ε_t, are stationary, then Assumption ST is satisfied.*

Proof. See Appendix A.1. □

To illustrate the role of the Stationarity restriction, we revisit the two-consumer, three-period example in Section 3.3.1, where we observe outcomes $\mathbf{Y}_1 = (0,0,1)$ and $\mathbf{Y}_2 = (0,1,1)$. Recall that eight of the 32 possible potential outcome paths are consistent with data. The observational equivalence condition requires that $\mathbb{P}[\mathbf{Y}_i = (0,0,1)] = P[\mu \in (0,(0,\cdot),(1,\cdot))] = 1/2$ and $\mathbb{P}[\mathbf{Y}_i = (0,1,1)] = P[\mu \in (0,(1,\cdot),(\cdot,1))] = 1/2$.

Assumption ST imposes the following additional restrictions,

$$P[\mu \in (\cdot,(0,0),(\cdot,\cdot))] = P[\mu \in (\cdot,(\cdot,\cdot),(0,0))] \tag{3.14}$$

$$P[\mu \in (\cdot,(0,1),(\cdot,\cdot))] = P[\mu \in (\cdot,(\cdot,\cdot),(0,1))] \tag{3.15}$$

$$P[\mu \in (\cdot,(1,0),(\cdot,\cdot))] = P[\mu \in (\cdot,(\cdot,\cdot),(1,0))] \tag{3.16}$$

$$P[\mu \in (\cdot,(1,1),(\cdot,\cdot))] = P[\mu \in (\cdot,(\cdot,\cdot),(1,1))] \tag{3.17}$$

While Assumption ST would reject the data if we only observed one of the two consumers, it adds identifying information in our panel that pools consumers. Starting with the right-hand side of (3.14), we have $P[\mu \in (\cdot,(\cdot,\cdot),(0,0))] = 0$ because this potential outcomes pattern is inconsitent with either $\mathbf{Y}_1$ or $\mathbf{Y}_2$. It follows that on the left-hand-side, which is only consistent with Y_1, we have $P[\mu \in (\cdot,(0,0),(\cdot,\cdot))] = 0$. Therefore, under Assumption ST, $\mathbf{Y}_1$ can only rationalized with $\mu \in (0,(0,1),(1,\cdot))$, giving us: $\mathbb{P}[\mathbf{Y}_i = (0,0,1)] = P[\mu \in (0,(0,1),(1,\cdot))] = 1/2$. In sum, Assumption ST rules out heterogeneity for $\mathbf{Y}_1$ in period $t = 1$.

Turning to (3.15), the left-hand-side potential outcomes pattern $(\cdot,(0,1),(\cdot,\cdot))$ is only consistent with $\mathbf{Y}_1$ and so $P[\mu \in (\cdot,(0,1),(\cdot,\cdot))] = P[\mu \in (0,(0,1),(1,\cdot))] = \mathbb{P}[\mathbf{Y}_i = (0,0,1)] = 1/2$, where the first equality comes from observational equivalence and the second equality comes from the previous step, where Assumption ST ruled out heterogeneity for $\mathbf{Y}_1$ in period $t=1$. By Assumption ST, the right-hand-side $P[\mu \in (\cdot,(\cdot,\cdot),(0,1))] = 1/2$. Since the right-hand-side is only consistent with $\mathbf{Y}_2$, it implies that $P[(0,(1,\cdot),(0,1))] = 1/2$ and so $\mathbf{Y}_2$ is only rationalized with $\mu \in (0,(1,\cdot),(0,1))$. Therefore, Assumption ST also rules out heterogeneity for $\mathbf{Y}_2$ in period $t=2$.

Restrictions (3.16) and (3.17), together with (3.14), (3.15), and observational equivalence, imply that $P[\mu \in (0,(0,1),(1,\tilde{u}))] = P[\mu \in (0,(1,\tilde{u}),(0,1))]$ for $\tilde{u} = 0,1$. The combination of the data and Assumption ST therefore generate the identified set

$$
\begin{aligned}
\mathscr{P}^{\star}_{ST} = \{P: \quad & \sum_{\mu \in \mathscr{U}} P(\mu) = 1, \quad P(\mu) \in [0,1] \\
& P[\mu \in (0,(0,1),(1,\cdot))] = \mathbb{P}[\mathbf{Y}_i = (0,0,1)] = \frac{1}{2} \\
& P[\mu \in (0,(1,\cdot),(0,1))] = \mathbb{P}[\mathbf{Y}_i = (0,1,1)] = \frac{1}{2} \\
& P[\mu = (0,(1,\tilde{u}),(0,1))] = P[\mu = (0,(0,1),(1,\tilde{u}))] \quad \text{for } = 0,1\}
\end{aligned}
$$

Furthermore, we have $HBL_1(P) = P[\mu \in (\cdot,(0,1),(\cdot,\cdot))] = P[\mu \in (\cdot,(0,1),(1,\cdot))] = 1/2$ and $HBL_2(P) = P[\mu \in (\cdot,(\cdot,\cdot),(0,1))] = P[\mu \in (\cdot,(1,\cdot)),(0,1)] = 1/2$ for all $P \in \mathscr{P}^{\star}_{ST}$. Therefore, $H\hat{B}L_1$ and $H\hat{B}L_2$ are point identified under stationarity in this example and 50% of the population exhibits HBL in each period.

Intuitively, Assumption ST sharpens the identified set by restricting the flexibility of population heterogeneity in the data, thereby bounding the degree of HBL from below. In the example above, Assumption ST excluded heterogeneity of the form $(\cdot,(0,0),(\cdot,\cdot))$ in $t=1$ as well, effectively bounding the HBL parameters above 0.

3.4.2 Monotone Treatment Selection (MTS)

Another potential restriction comes from the assumption that self-selection into loyalty for a brand j (e.g., $Y_{i(t-1)} = j$) contains identifying information about a consumer's latent preference for j. We use Torgovitsky (2019)'s Monotonte Treatment Selection restriction:

Assumption MTS: For $\forall y, \tilde{y} \in \{1,...,J\}$ and all $t \geq 2$ s.t. $\mathbb{P}[Y_{i(t-1)} = j | Y_{i(t-2)} = \tilde{y}] \in (0,1)$,

$$\mathbb{P}_P[U_{it}(y) = j | Y_{i(t-1)} = j, Y_{i(t-2)} = \tilde{y}] \geq \mathbb{P}_P[U_{it}(y) = j | Y_{i(t-1)} \neq j, Y_{i(t-2)} = \tilde{y}] \tag{3.18}$$

Assumption MTS implies that a consumer who purchased brand j in period $t-1$ is more likely to choose brand j in all potential loyalty state during period t than some other consumer who does not choose j in period $t-1$. The additional conditioning on $Y_{i(t-2)}$ ensures that $Y_{i(t-1)}$ are comparable. In other words, one could interpret the MTS assumption as saying consumers who self selected into loyalty have a higher latent propensity to purchase the same brand in the following period than those who didn't select into loyalty. In the RUM context, such latent propensity to purchase may come from a persistent unobserved preference for the brand, or from positively serial correlated utility shocks, as an example.

The following proposition establishes that weakly positive serial-dependence in the utility shocks is sufficient for the potential ouctomes to satisfy MTS under the static version of the RUM in Section 2 (i.e., with consumer discount factor $\beta = 0$) when choice-specific utility takes the canonical form $u_j(p,l) = \gamma_j - \alpha p_j + \lambda \mathbb{I}_{\{l=j\}}$, prices are time-invariant and the choice set contains two brands (i.e. $|\mathscr{J}| = 2$).

Proposition 3. *Suppose that every $P \in \mathscr{P}^\dagger$ is consistent with $U_{it}(y)$ being generated through the RUM (3.3) with $\mathscr{J} = \{1,2\}$, $u_j(p,l) = \gamma_j - \alpha p_j + \lambda \mathbb{I}_{\{l=j\}}$, and time invariant prices $\mathbf{p}_t = \mathbf{p}_{t'}$ for all t,t'. If the utility shocks ε_{jt} and $\varepsilon_{j(t+1)}$ have weakly positive covariance for all j,t, $\varepsilon_{jt} \perp \varepsilon_{kt}$ for all j,k,t, and $\varepsilon_t \perp p$ for all t, then Assumption MTS is satisfied.*

Proof. See Appendix A.1. □

Proposition 4. *Suppose that every $P \in \mathscr{P}^\dagger$ is consistent with $U_{it}(y)$ being generated through the RUM (3.3) with time invariant prices $\mathbf{p}_t = \mathbf{p}, \forall t$. If the choice-specific probabilities $\sigma_j(P,y)$ and $\sigma_j(P,\tilde{y})$ have weakly positive covariance for all $y, \tilde{y} \in \mathscr{J}$ then Assumption MTS is satisfied.*

Proof. See Appendix A.1. □

We have not worked out the extension of Proposition 3 to the version of the DDC that includes potentially serially-dependent covariates, such as prices. In that case, MTS in the RUM would also require restrictions over the distribution of prices.

To illustrate the role of Assumption MTS, we again revisit the two-consumer, three-period example in Section 3.3.1, where the observational equivalence condition alone gives us: $H\hat{B}L_2 = P[\mu \in (\cdot,(\cdot,\cdot),(0,1))] = P[\mu \in (0,(1,\cdot),(0,1))] \leq \mathbb{P}[\mathbf{Y}_i = (0,1,1)] = 1/2$. Assumption MTS imposes the following additional restriction:

$$P[U_{i2}(0) = 1|Y_{i1} = 1, Y_{i0} = 0] \geq P[U_{i2}(0) = 1|Y_{i1} = 0, Y_{i0} = 0] \tag{3.19}$$

The right hand side of (3.19) is consistent with $\mathbf{Y}_1$ and occurs with probability 1: $\mathbb{P}[Y_{i2} = 1|Y_{i1} = 0, Y_{i0} = 0] = 1$. The inequality implies that the left hand side must give $P[U_{i2}(0) = 1|Y_{i1} = 1, Y_{i0} = 0] = 1$. Since the conditioning event on the left-hand-side can only be consistent with $\mathbf{Y}_2$, Assumption MTS rules out HBL for consumer 2 in $t = 2$: $H\hat{B}L_2 = P[\mu \in (0,(1,\cdot),(0,1))] = 0$. Therefore, Assumption MTS reduces the upper bound of $H\hat{B}L_2$ from 1/2 to 0.

Intuitively, Assumption MTS sharpens the identified set by restricting the sources of serial-dependence in the data which can be confounded with HBL. The restriction bounds the degree of HBL from above.

3.4.3 Monotone Treatment Response

A natural candidate restriction is directly on sign of the state dependence in brand choices. We could explicitly rule out negative state dependence, i.e. variety-seeking, through the following restriction:

Assumption MTR: For $\forall k \neq j$ and for all t,

$$\mathbb{P}_P[\mathbb{1}\{U_{it}(j) = j\} \geq \mathbb{1}\{U_{it}(k) = j\}] = 1 \tag{3.20}$$

This assumption captures the idea that being loyal to a brand weakly improves the utility of that brand more than the utility of any other brand. Assumption MTR imposes uniformity over the evolution of choices in the sense that a change in the loyalty state from k to j should increase everyone's propensity to choose j.

The following proposition establishes a sufficient condition for the RUM in section 2 to imply MTR. In the usual linear-index version of the RUM, Assumption MTR holds if the loyalty coefficient is non-negative. More generally, this assumption is implied by a separable additive utility function where a change in state from k to j increases the utility of option j relative to all other options (Heckman and Vytlacil, 2007; Heckman et al., 2008; Lee and Salanié, 2018; Irace, 2018).

Proposition 5. *Suppose that every $P \in \mathscr{P}^{\dagger}$ is consistent with $U_{it}(y)$ being generated through (3.3) by an RUM model. If $u_j(\mathbf{p}, l = j) - u_j(\mathbf{p}, l = k) \geq 0$ and $u_l(\mathbf{p}, l = j) - u_l(\mathbf{p}, l = k) \leq 0$ at all $\mathbf{p}$ for all $k, l \neq j \in \mathscr{J}$, then Assumption MTR is satisfied.*

Proof. See Appendix A.1. □

While MTR may not always sharpen the bounds of the HBL parameters, it reduces the computational burden by eliminating potential outcomes paths. In the two-consumer, three-period example of section 3.3.1, Assumption MTR reduces the number of adimissible potential outcome paths from 8 to 4, so that $\mathbf{Y}_1$ is only consistent with $\mu \in (0, (0, \cdot), (1, 1))$ and $\mathbf{Y}_2$ is only consistent with $\mu \in (0, (1, 1), (\cdot, 1))$.

3.4.4 Omitted Marketing Variables

One advantage of the DPO model (3.1) is that it does not impose any structure on the manner in which marketing variables, such as prices or advertising affect choices. However, a potential concern is that the presence of unmodeled, demand-shifting marketing variables might violate some of our identifying assumptions. For instance, if prices are non-stationary, then to the extent they change demand, the potential outcomes may also be non-stationary, violating our ST restriction. Incorrectly imposing ST could spuriously identify a non-zero lower bound on the HBL parameter.[5] Therefore, we must implicitly assume that prices (and other marketing variables) and the resulting potential outcomes are stationary so that our model is not mis-specified under ST. For instance, ST would not be violated if unmodeled prices follow a Markov process.[6] The implications of unmodeled marketing variables for MTS are more complicated. Alternatively, one could in principle re-define the HBL parameter to be conditional on prices (or other marketing variables).

[5] Recall that non-stationarity in choices identifies HBL under the ST restriction.

[6] Eichenbaum, Jaimovich and Rebelo (2011) find that CPG prices follow Markov process whereby prices alternate between *regular* and *discount* prices, and Dekimpe, Hanssens and Silva-Risso (2011) find that CPG prices follow a linear, first-order Markov process.

Concerns about omitted marketing variables and spurious HBL are not unique to the DPO framework. In the stochastic brand choice literature, one needs to make the much stronger assumption that marketing variables do not shift the choice probabilities under the null multinomial process used for the runs tests and t-tests. Under the DPO, we only need to ensure that omitted marketing variables do not invalidating our identifying restrictions on the potential outcomes distribution.

3.5 Computing the Identified Set of HBL Parameters

Every probability mass function P in the identified set $\mathscr{P}^\star$ produces an HBL instance that is consistent with data and prior assumptions ($\rho(P) \geq 0$, including Assumptions ST, MTS and MTR). To get the identified set of $\hat{HBL}$, we use the linear programming approach from Torgovitsky (2019). After extending the objective functions and contraints to the multinomial choice context as we discussed above, we can compute the identified set of $\hat{HBL}$ as a closed interval $\Theta^\star \equiv [\theta^\star_{lb}, \theta^\star_{ub}]$ by solving two optimization problems,

$$\theta^\star_{lb} \equiv \min_{\{P(\mu)\in[0,1]:\mu\in\mathscr{U}\}} \theta(P) \quad s.t.\ \rho(P) \geq 0,\ (3.7),\ \text{and } (3.9) \tag{3.21}$$

$$\theta^\star_{ub} \equiv \max_{\{P(\mu)\in[0,1]:\mu\in\mathscr{U}\}} \theta(P) \quad s.t.\ \rho(P) \geq 0,\ (3.7),\ \text{and } (3.9) \tag{3.22}$$

4 A Test of Forward Looking Behavior in Brand Loyalty

The DPO nests a broad class of RUMs, including myopic consumers who fully discount the future ($\beta = 0$) and passively respond to their brand habits, and forward-looking consumers ($\beta > 0$) who consciously plan their future brand loyalty. The extant empirical literature using RUMs to study habitual brand loyalty typically assumes myopic consumer behavior.[7] We propose additional parametric restrictions under which the discount factor, β, can be jointly identified with the consumer's deterministic flow utilities, $u_j(s), \forall j \in \mathscr{D}$. Our identification results are constructive and immediately

[7] See for instance Guadagni and Little (1983), Keane (1997), Seetharaman et al. (1999), and Dubé et al. (2010). One exception is Chintagunta et al. (2001) who estimate the reduced-form of the DDC in 2.3 without separately identifying β and preferences.

suggest a semi-parametric minimum-distance estimator of the discount factor and test for forward-looking.

4.1 Identification of the Discount Factor

The extant literature has established that the discount factor β is not identified in discrete-choice RUMs without additional restrictions (Rust, 1994; Magnac and Thesmar, 2002). We now show that the DDC in section 2 has built-in exclusion restrictions of the form discussed in (Abbring and Daljord, 2020) that resolve this non-identification problem.

Following Hotz and Miller (1993), the inversion of (2.2) identifies the choice-specific value contrasts:

$$\begin{aligned} \ln\left(\frac{\sigma_j(s)}{\sigma_0(s)}\right) &= v_j(s) - v_0(s) \\ &= u_j(s) + \beta[\mathbf{F}_j(s) - \mathbf{F}_0(s)][\mathbf{I} - \beta \mathbf{F}_0]^{-1}\mathbf{m} \end{aligned} \tag{4.1}$$

where $\mathbf{F}_j = \left[F_0(s_1)', ..., F_0(s_L)'\right]'$ are the consumer's beliefs over future states conditional on choice $j \in \mathscr{D}$. However, (4.1) reveals that even if beliefs $\mathbf{F}$ are observed, an additional moment is needed to disentangle the utilities, $u(s)_j$, and the discount factor, β.

In appendix A.2, we show that the canonical differentiated-products, discrete-choice RUM has a built-in set of exclusion restrictions due to the fact that the current-period utility only depends on the price of the chosen alternative, p_j, and not on the prices of rival products, $\mathbf{p}_{-j}$:

$$u_j\left(p_j, \mathbf{p}_{-j}, \ell\right) = u_j\left(p_j, ., \ell\right).$$

Let $\mathscr{S}^i d$ denote the set of pairs of states that generate excluded rival price variation:

$$\mathscr{S}^{id} = \left\{ \left\{s, s'\right\} \in \mathscr{S}^2 : p_j = p_j',\ \mathbf{p_{-j}} \neq \mathbf{p'_{-j}},\ \text{and } \ell = \ell' = l, l \neq j, \text{ for } j \in \mathscr{D}_{-0} \right\}. \tag{4.2}$$

If we assume that prices are exogenous, $\mathbb{E}(\varepsilon_{jt}|p_{jt}) = 0$, we can construct a new set of moment conditions by differencing the Hotz-Miller conditions (4.1) for each pair of states $(s,s')\in\mathscr{S}^{id}$:

$$\ln\left(\frac{\sigma_j(s)}{\sigma_0(s)}\right)-\ln\left(\frac{\sigma_j(s')}{\sigma_0(s')}\right)=\beta[\mathbf{F}_j(s)-\mathbf{F}_0(s)-\mathbf{F}_j(s')+\mathbf{F}_0(s')][\mathbf{I}-\beta\mathbf{F}_0]^{-1}\mathbf{m}. \tag{4.3}$$

The following proposition, based on Theorem 1 in Abbring and Daljord (2020), characterizes the identified set for β based on the moments (4.3).

Proposition 6. *Suppose that we observe pairs of states as in (4.2) and that either the left hand side of (4.3) is nonzero (that is, $\sigma_j(s)/\sigma_0(s)\neq\sigma_j(s')/\sigma_0(s')$) or $[F_j(s)-F_0(s)-F_j(s')+F_0(s')][I-\beta F_0]^{-1}m\neq 0$. Then the identified set of discount factors contains at most L elements.*

Proof. See Abbring and Daljord (2020). □

In practice, the rank condition in theorem 6 does not guarantee that β is point identified. However, Abbring and Daljord (2020) conjecture that when multiple exclusion restrictions are available, as in our empirical applications below, β will be point identified. More important, conditional on β, we can point-identify the utilities from the Hotz-Miller conditions (4.1).

4.2 A Semi-Parametric Minimum Distance Estimator for the Discount Factor

We construct the empirical analog of the moments (4.3) for each observed pair $(s',s'')\in S^{id}$

$$\mathbf{g}(\beta)=\ln\left(\frac{\sigma_a(s')}{\sigma_0(s')}\right)-\ln\left(\frac{\sigma_a(s'')}{\sigma_0(s'')}\right)-\beta\left[F_a(s')-F_0(s')-F_a(s'')+F_0(s'')\right][I-\beta F_0]^{-1}m. \tag{4.4}$$

We then define the miminum distance estimator

$$\beta^{MD}=\underset{\beta}{\operatorname{argmin}}\,\mathbf{g}(\beta)'\mathbf{W}\mathbf{g}(\beta) \tag{4.5}$$

where $\mathbf{W}$ is an (R) weight matrix and $\mathbf{g}(\beta)$ is the $(R\times 1)$ vector of moments. Inference follows directly from Newey and McFadden (1994).

5 Empirical Application: CPG Demand

5.1 Data

We conduct a case study of demand for consumer brands in the CPG industry using the IRI Academic Dataset.[8] The data comprise a shopping panel for 15,079 households spanning 12 years between 2001 and 2012. In addition, the data include weekly price and sales data at the UPC level for chain grocery and drug stores in those markets. Our analysis focus on the period from 2004 to 2007, since this data period provides the largest sample with qualifying purchases and that a longer period is more likely to contain non-stationary changes in the purchase environment that may invalidate the stationarity assumption in the DPO model.

For our analysis, we select CPG product categories that have oligolisitic market structure, that are not susceptible to stockpiling and addictive behavior, and with sufficient number of observations. For each category, we focus on the top selling product form and package size, and construct purchase panel for top 2-6 brands in terms of purchase incidences. Horizontally differentiated brand variants are aggregated if their price correlation is high. Within each category, we retain those households that satisfy minimal requirements for reporting, make in-category purchases only at stores with complete price records for all brands in the choice set, only purchase one or less brand from the choice set each week, and make at least three in-category purchases in the data period. We keep all trips made by these households that took place in stores with a complete price record. Outside option is defined as a trip without purchasing brands in the choice set.

After applying our screening criteria, we obtain panel data for the following 8 categories: coffee, deodoreant, mayonnaise, margarine, peanut butter, spaghetti sauce, tooth brushes, and yogurt. Table 1 provides summary statistics of each of the 8 estimation samples.

5.2 Evidence of Brand Loyalty

To implement the DPO model on brand choice data in CPG categories, we adopt several measures to reduce the computational burden of the DPO problem. The dimensionality of the DPO problem grows exponentially in both J and T, for there are J^{JT+1} variables each being a probability of a

[8] see Bronnenberg et al. (2008) and https://www.iriworldwide.com/en-us/solutions/academic-data-set

category	brands	households	trips	purchases	brands
coffee	5	870	139.7	8.7	MAXWELL Sm, FOLGERS Sm, FOLGERS Lg
deodorant	6	1143	197.6	5.1	MENNEN, OLDSPICE, DOVE, DEGREE, RIG
mayonnaise	3	2109	173	7	HELLMANNS, PL, KRAFT
margarine	4	1241	166.5	11	SMARTBALANCE, ICBINB, SHEDDSCOUNT
peanut butter	4	4143	170.1	8.9	SKIPPY, JIF, PL, PETERPAN
spaghetti sauce	4	1899	185	11.5	RAGU, PREGO, FRANCESCORINALDI, HUN
toothbrushes	2	505	241.5	4.5	COLGATE, ORALB
yogurt	3	1457	168.4	13.3	YOPLAIT, DANNON, COLOMBO

This table summarizes the estimation samples drawn from the IRI Academic Dataset. The number of trips and number of purchases are computed as cross-household means.

Table 1: Descriptive Statistics

potential outcomes path. For a modest size of problem with $J = 3$ and $T = 4$, there are more than 1.5 million variables and at least 241 constraints for observational equivalence. To control the size of J, we drop no-purchases, create a multinomial-choice estimation sample with top two brands in each category and consolidate the remainder brands in to a composite good ($J = 3$), as well as a binary choice estimation sample with top one brand and consolidate the remainder brands ($J = 2$). To control the size of T, we keep the last six purchases of each household for DPO estimation. With the multinomial estimation sample, we further adopt a dimensionality reduction technique introduced in Torgovitsky (2019) to construct shorter DPO models (ML=3) and use a collection these shorter models to relate potential outcomes to observed brand choices.

In theory, one could also include time-varying covariates X, such as prices, into the DPO model by modifying primitives $P : \mathscr{U} \times \mathscr{X}$. However, it will further exacerbate the dimensionality issue, for that now we will have $J^{JT+1} \times |\mathscr{X}|$ variables to solve for. Therefore, we suppress prices in the DPO model and focus on loyalty states only.

We first report the share of the consumer choice occasions that exhibit HBL for the most popular brand. Table 2 reports the cross-time average HBL estimated from the multinomial-choice estimation sample ($J = 3$), and Table 3 reports HBL bounds and confidence regions estimated from a binary-choice estimation sample. The target parameter is $HBL_j \equiv \frac{1}{T}\sum_{t=1}^{T} HBL_{jt}(P)$, where j denotes the most popular brand in each category. We impose either no identifying assumptions, or a set of assumptions including $ST(1)$, MTS and MTR. The estimated bounds and the 95% confidence regions are constructed using the procedure of Torgovitsky (2019) and Chernozhukov et al. (2015). Figure 2 and Figure 3 visualize the estimates.

For a marketer, the more interesting question may be to assess the extent to which its current consumer share is due to HBL. So we also report HBL conditional on being a current purchaser of the most popular brand, $HBL|Y_t = j$. Table 4 and Figure 4 report the bounds for the cross-time average of $HBL|Y_t = j$. XXXX DISCUSSION XXXX

Finally, to decompose the persistence in brand choice for the most popular brand, we also report HBL conditional on both currently choosing the most popular brand and also choosing it the previous trip, $HBL|Y_t = jandY_{t-1} = j$. Table 5 and Figure 5 report the bounds for the cross-time average of $HBL|Y_t = jandY_{t-1} = j$ under several alternative sets of restrictions. XXXX DISCUSSION XXXX

category	brand	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$	$95\%CR_{lb}$	$95\%CR_{ub}$
coffee	MAXWELL$_S$	0.070	0.395	0.029	0.489
deodorant	MENNEN	0.071	0.513	0.000	0.617
mayonnaise	HELLMANNS	0.042	0.579	0.023	0.671
margarine	SMART BALANCE	0.037	0.574	0.000	0.706
peanut butter	SKIPPY	0.056	0.485	0.046	0.535
spaghetti sauce	RAGU	0.029	0.493	0.000	0.536
toothbrush	COLGATE	0.162	0.242	0.053	0.423
yogurt	YOPLAIT	0.038	0.526	0.000	0.603

This table summarizes the bounds on HBL using the estimation samples drawn from the IRI Academic Dataset. We report bounds for the cross-time average value of HBL_j where j is the top-selling brand in each category. Estimation is based on a multinomial DPO model with $J = 3$ (except for the toothbrush category where $J = 2$), $T = 5$ which stitches together sequences of length $ML = 3$ using the dimension-reduction technique in Torgovitsky (2019). The choice sets consist of each of the top-two selling brands and a composite brand representing all other choice alternatives. $\hat{HBL}_{lb}$ and $\hat{HBL}_{ub}$ incidates the estimated bounds. $95\%CR_{lb}$ and $95\%CR_{ub}$ indicate the 95% confidence region constructed using the CNS method discussed in Torgovitsky (2019) and Chernozhukov et al. (2015), implemented using 30 bootstrap draws and a tuning parameter of $\tau_n = 0.8$.

Table 2: Habitual Brand Loyalty in IRI Data - Multinomial DPO

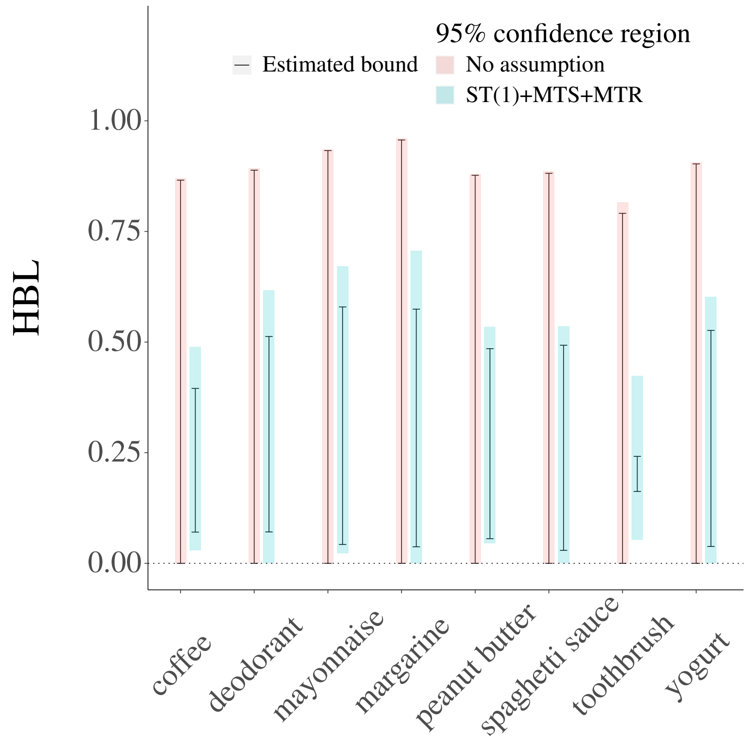

Figure 2: Share of population with HBL - Multinomial DPO

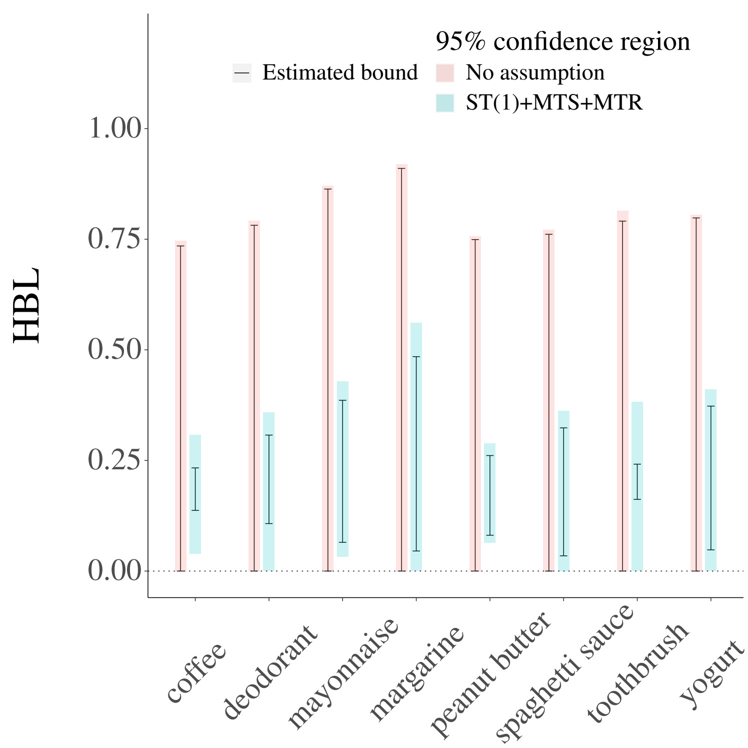

Figure 3: Share of population with HBL - Binary DPO

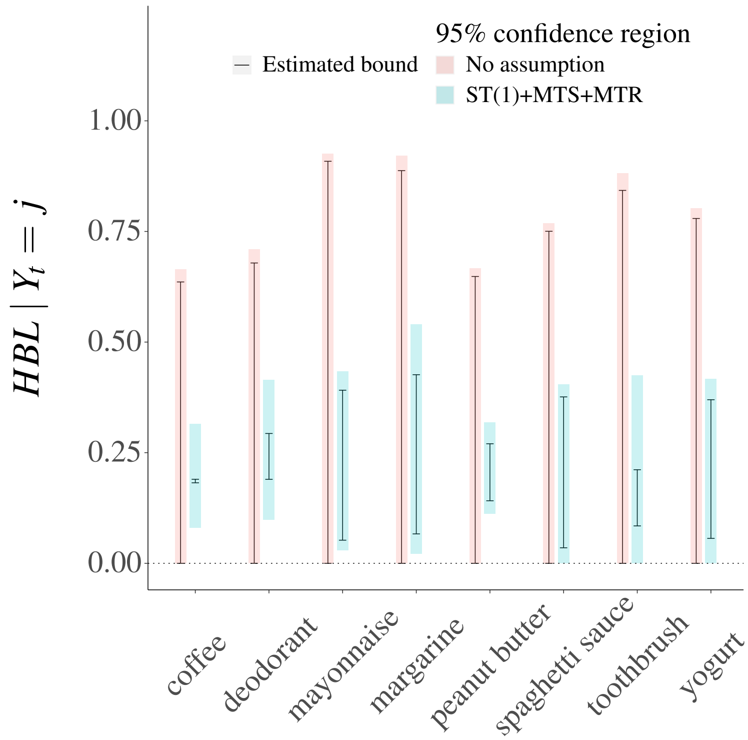

Figure 4: Share of Current Buyers with HBL - Binary DPO

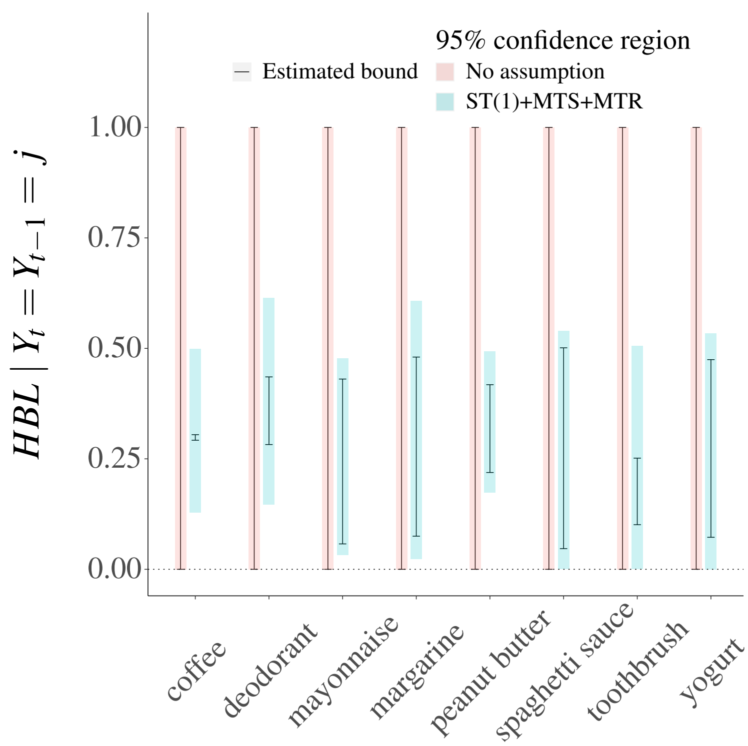

Figure 5: Share of Current Repeat Buyers with HBL - Binary DPO

category	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$	$95\%CR_{lb}$	$95\%CR_{ub}$
coffee	0.165	0.275	0.000	0.346
deodorant	0.201	0.306	0.000	0.438
mayonnaise	0.063	0.358	0.008	0.484
margarine	0.057	0.485	0.000	0.594
peanut butter	0.108	0.261	0.032	0.326
spaghetti sauce	0.079	0.324	0.000	0.381
toothbrush	0.130	0.320	0.000	0.503
yogurt	0.107	0.373	0.000	0.456

This table summarizes the bounds on HBL using the estimation samples drawn from the IRI Academic Dataset. We report bounds for the cross-time average value of HBL_j where j is the top-selling brand in each category. Estimation is based on a binomial DPO model with $J = 2$ and $T = 5$. The choice sets consist of the top-selling brand and a composite brand representing all other choice alternatives. $\hat{HBL}_{lb}$ and $\hat{HBL}_{ub}$ incidate the estimated bounds. $95\%CR_{lb}$ and $95\%CR_{ub}$ indicate the 95% confidence region constructed using the CNS method discussed in Torgovitsky (2019) and Chernozhukov et al. (2015), implemented using 30 bootstrap draws and a tuning parameter of $\tau_n = 0.8$.

Table 3: Habitual Brand Loyalty in IRI Data - Binomial DPO

5.3 Evidence of Forward Looking in Brand Loyalty

For our empirical application of the DDC, we need to specify the transition process for the loyalty state ℓ. We follow the convention in the literature and assume

$$\ell_t = \begin{cases} j & \text{if } d_{t-1} = j \in \mathscr{D}_{-0}, \\ \ell_t & \text{if } d_{t-1} = 0. \end{cases} \tag{5.1}$$

We also modify the consumer's decision problem to accomodate stochastic timing of trips. We use the cross-household mean probability of making a trip each week π. We can then modify our moments (4.3) as follows:

$$\ln\left(\frac{\sigma_j(s)}{\sigma_0(s)}\right) - \ln\left(\frac{\sigma_j(s')}{\sigma_0(s')}\right) = \beta\pi\left(\mathbf{F}_j(s)\mathbf{B_j} - \mathbf{F}_0(s)\mathbf{B_0} - \mathbf{F}_j(s')\mathbf{B_j} + \mathbf{F}_0(s')\mathbf{B_0}\right)\left[\mathbf{I} - \beta\pi\mathbf{F}_0\mathbf{B_0}\right]^{-1}\mathbf{m}$$

where $\mathbf{B}_j = \left[\mathbf{I} - \beta\,(1-\pi)\,\mathbf{F}_j\right]^{-1}$.

To control for unobserved heterogeneity, we use the group fixed-effects approach of Bonhomme et al. (2021). For details, please see Appendix A.4.

We start with our semi-parametric estimates of β using the minimum distance estimator (4.5). The first two columns of Table 6 report the estimated population mean and standard error, respectively,

category	brand	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$	$95\%CR_{lb}$	$95\%CR_{ub}$
coffee	MAXWELL_S	0.182	0.190	0.080	0.314
deodorant	MENNEN	0.190	0.293	0.099	0.415
mayonnaise	HELLMANNS	0.052	0.391	0.030	0.434
margarine	SMART BALANCE	0.066	0.426	0.021	0.540
peanut butter	SKIPPY	0.141	0.270	0.112	0.319
spaghetti sauce	RAGU	0.035	0.376	0.000	0.405
toothbrush	COLGATE	0.085	0.211	0.000	0.425
yogurt	YOPLAIT	0.056	0.370	0.000	0.417

This table summarizes the bounds on HBL conditional on purchasing the top-selling brand using the estimation samples drawn from the IRI Academic Dataset. We report bounds for the cross-time average value of $HBL_j|Y_t = j$ where j is the top-selling brand in each category. Estimation is based on a binomial DPO model with $J = 2$ and $T = 5$. The choice sets consist of the top-selling brand and a composite brand representing all other choice alternatives. $\hat{HBL}_{lb}$ and $\hat{HBL}_{ub}$ incidate the estimated bounds. $95\%CR_{lb}$ and $95\%CR_{ub}$ indicate the 95% confidence region constructed using the CNS method discussed in Torgovitsky (2019) and Chernozhukov et al. (2015), implemented using 30 bootstrap draws and a tuning parameter of $\tau_n = 0.8$.

Table 4: Habitual Brand Loyalty Among Current Buyers - Binomial DPO

for each category. With the exception of spaghetti sauce, all of the point estimates are significantly different from zero implying that consumers are indeed forward-looking. However, our point estimates range from as low as 0.41 to very close to one. Given that we model a weekly choice, these results suggest that most categories have very impatient consumers.A discount factor of $\beta = 0.9$, for instance, implies an annual interest rate of 238%. Consumer impatience has also been documented in the extant literature (e.g., Frederick et al. (2002); Dubé et al. (2014)).

We now turn to the parametric estimates from the DDC model. For the empirical application, we assume:

$$u_j(s_t) = \gamma_j - \alpha p_{jt} + \lambda \mathbb{I}_{\{\ell_t = j\}},\ \forall j \in \mathscr{D}_{-0}.$$

We use an MPEC implementation of the maximum likelihood estimator. For details, please see Appendix A.3. To control for unobserved heterogeneity, we once again use the group fixed-effects approach of Bonhomme et al. (2021). For details, please see Appendix A.4.

The third and fourth columns of Table 6 report the discount factor estimates from the DDC. All of our estimates are statistically significant (margarine only at the 10% level). Using the likelihood ratio test, we reject the myopic model (with $\beta = 0$) in favor of the DDC with freely-varying β at the 5% significance level for all the categories except coffee (results reported in Table A19 in Appendix

category	brand	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$	$95\%CR_{lb}$	$95\%CR_{ub}$
coffee	MAXWELL_S	0.292	0.305	0.128	0.499
deodorant	MENNEN	0.282	0.435	0.147	0.615
mayonnaise	HELLMANNS	0.058	0.431	0.033	0.478
margarine	SMART BALANCE	0.075	0.480	0.024	0.608
peanut butter	SKIPPY	0.219	0.418	0.174	0.493
spaghetti sauce	RAGU	0.047	0.501	0.000	0.540
toothbrush	COLGATE	0.101	0.252	0.000	0.506
yogurt	YOPLAIT	0.072	0.475	0.000	0.534

This table summarizes the bounds on HBL conditional on repeat-purchasing the top-selling brand using the estimation samples drawn from the IRI Academic Dataset. We report bounds for the cross-time average value of $HBL_j|Y_t = jandY_{t-j} = j$ where j is the top-selling brand in each category. Estimation is based on a binomial DPO model with $J = 2$ and $T = 5$. The choice sets consist of the top-selling brand and a composite brand representing all other choice alternatives. $\hat{HBL}_{lb}$ and $\hat{HBL}_{ub}$ incidate the estimated bounds. $95\%CR_{lb}$ and $95\%CR_{ub}$ indicate the 95% confidence region constructed using the CNS method discussed in Torgovitsky (2019) and Chernozhukov et al. (2015), implemented using 30 bootstrap draws and a tuning parameter of $\tau_n = 0.8$.

Table 5: Habitual Brand Loyalty Among Current Repeat Buyers - Binomial DPO

A.8). Once again, we see a wide range of values, implying a fair amount of consumer impatience. However, our DDC results do not always line up exactly with the semi-parametric estimates. We reject the hypothesis that the semi-parametric estimate of β equals the DDC estimate in 3 of our 8 categories (deodorant, mayonnaise, and yogurt).[9]

We report the full set of taste coefficient estimates for the static RUM and DDC specifications, respectively, in Tables A20 to A27 in Appendix A.8. As in the extant literature, we find large coefficient estimates for brand loyalty, λ, even after controlling for unobserved heterogeneity. In most of the categories, the expected, dollar-denominated loyalty premium is over \$1 ($-\frac{\lambda}{\alpha}$). Surprisingly, we find almost no correlation between the estimated discount factor, β, and the other taste coefficients in all of the categories. Consequently, the taste coefficients for the static and dynamic models are quite similar in magnitude.

To compare our parametric results herein to the non-parametric results from section **??**, we compute the parametirc analog of the statistic **HBL**, as in (3.6):

$$HBL_j = \mathbb{P}_\varepsilon[\exists k \neq j \in \mathscr{J} \text{ s.t. } j = \underset{l \in \mathscr{J}}{\text{argmax}}\{v_{il}(\mathbf{p}_t, j) + \varepsilon_{ilt}\}, \text{ and } j \neq \underset{l \in \mathscr{J}}{\text{argmax}}\{v_{il}(\mathbf{p}_t, k) + \varepsilon_{ilt}\}, \forall i, t].$$

[9]We test this hypothesis using the z-statistic: $\frac{\beta^{MD} - \beta^{DDC}}{\sqrt{se^2_{MD} + se^2_{DDC}}}$ **??**

category	Min. Dist.	Min. Dist. (s.e.)	DDC	DDC (s.e.)
coffee	0.680	0.156	0.824	0.148
deodorant	0.999	0.066	0.998	0.011
mayonnaise	0.739	0.121	0.999	0.002
margarine	0.638	0.110	0.704	0.205
peanut butter	0.999	0.005	0.993	0.021
spaghetti sauce	0.503	0.159	0.914	0.096
toothbrushes	0.537	0.203	0.550	0.175
yogurt	0.999	0.115	0.474	0.158

This table summarizes the discount factor estimates from the estimation samples drawn from the IRI Academic Dataset. "Min. Dist." indicates the semi-parameters, minimum distance estimator. "DDC" indicates the dynamic discrete-choice model estimates. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table 6: The Discount Factor β

See Appendix A.5 for details. Table 7 contains the results for the leading brand in each category. XXXX

category	brand	$\hat{HBL}_{MNL}$	$\hat{HBL}_{MNL}$ (s.e.)	$\hat{HBL}_{DDC}$	$\hat{HBL}_{DDC}$ (s.e.)
coffee	MAXWELL Sm	0.517	0.051	0.517	0.051
deodorant	MENNEN	0.650	0.022	0.667	0.067
mayonnaise	HELLMANNS	0.719	0.035	0.711	0.077
margarine	ICBINB	0.889	0.019	0.894	0.030
peanut butter	SKIPPY	0.605	0.015	0.617	0.016
spaghetti sauce	RAGU	0.630	0.029	0.636	0.032
toothbrushes	COLGATE	0.450	0.076	0.420	0.080
yogurt	YOPLAIT	0.741	0.026	0.743	0.026

This table summarizes the HBL statistics corresponding to the DDC estimates from the IRI Academic Dataset. HBL computed for the top brand in each of the categories. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table 7: The Extent of Habitual Brand Loyalty (HBL): DDC evidence

In sum, we find strong evidence for forward-looking consumer behavior across product categories. However, consumers appear to be more mypoic than would be implied by the real rate of interest. We explore the implications of this behavior in the next section.

6 Policy Implications

To asses the implications of consumer forward-looking behavior, we compare the short-run and long-run elasticity of demand. For each category, we simulate the long-term effect of a 10% increase in the price of the top-share brand in each state. We then compue the corresponding expected elasticities in equilibrium, using the steady-state probabilities of each state. For the short-run elasticities, we use the parameter estimates from the static version of the RUM with $\beta = 0$. For the long-run elasticities, we use the DDC estimates, which allow β to vary freely in $[0,1]$. For technical details of the simulation of elasticities, see Appendix A.6.

We report the results in Tables A1 to A8 in Appendix A.6. Even though the taste coefficients are fairly similar in the static RUM and DDC specifications, we nevertheless observe striking differences in the elasticities. The long-run elasticities are all considerably larger than the short-run elasticities, indicating that consumer demand is much more price-sensitive when consumers *plan* their future brand habits. The difference between the long-run and short-run own-price elasticities vary from as low as 2%, in margarine, to 35.6%, in peanut butter. Turning to the cross-price elasticities between brands, the long-run valuess are mostly hundreds or even thousands of percent higher than the short-run values (coffee is the exception with cross-elasticity differences of about 65%). However, perhaps as expected, we see smaller impact on the outside good. Intuitively, if consumers plan their future brand habits, then the "cost" of being loyal to the "wrong" brand is higher, making consumers more sensitive to permanent price changes.

In sum, the static model appears to under-estimate the price-sensitivity of demand in response to permanent price changes. In particular, the static model drastically under-estimates the substitution between brands in the long term.

7 Conclusions

We have documented non-parametric evidence of brand loyalty across several CPG product categories whereby a consumer's past brand choices have a causal effect on her current choice, similar to a "switching cost." We have also documented semi-parametric evidence that consumers make forward-looking purchase decisions and consciously *plan* their future loyalty habits. However, consumers appear to be more impatient than would be implied by the real rate of interest. Finally. using a parametric dynamic discrete-choice model of demand, we show that the long-run price elasticities of demand are considerably larger than would be implied by a static model with the deterministic restriction $\beta = 0$. In sum, the combination of brand loyalty and forward-looking behavior likely have a material impact on demand and market prices.

References

Jaap H Abbring and Øystein Daljord. Identifying the discount factor in dynamic discrete choice models. *Quantitative Economics*, 11(2):471–501, 2020.

Greg M. Allenby and Peter E. Rossi. Marketing models of consumer heterogeneity. *Journal of Econometrics*, 26(1-2):57–78, 1998.

Joe S. Bain. *Barriers to New Competition*. Cambridge: Harvard University Press, 1956.

Frank M. Bass, Moshe M. Givon, Manohar U. Kalwani, David Reibstein, and Gordon P. Wright. An investigation into the order of the brand choice process. *Marketing Science*, 3(4):267–287, 1984.

s. Bonhomme, T. Lamadon, and E. Manresa. Discretizing unobserved heterogeneity. Technical report, University of Chicago, 2021.

Bart J Bronnenberg, Michael W Kruger, and Carl F Mela. Database paper—the iri marketing data set. *Marketing science*, 27(4):745–748, 2008.

G. H. Brown. Brand loyalty-fact or fiction. *The Trademark Reporter*, 43:251–258, 1953.

Victor Chernozhukov, Whitney K Newey, and Andres Santos. Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*, 2015.

Pradeep Chintagunta, Ekaterini Kyriazidou, and Josef Perktold. Panel data analysis of household brand choices. *Journal of Econometric*, 103:111–153, 2001.

Carol Corrado, John Haltiwanger, and Daniel Sichel. Introduction to" measuring capital in the new economy". In *Measuring Capital in the New Economy*, pages 1–10. University of Chicago Press, 2005.

Ross M Cunningham. Brand loyalty-what, where, how much. *Harvard business review*, 34(1): 116–128, 1956.

Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445, 2010.

Jean-Pierre Dubé, Jeremy T. Fox, and Che-Lin Su. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267, 2012.

Jean-Pierre Dubé, Günter J. Hitsch, and Pranav Jindal. The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. *Quantitative Marketing and Economics*, 12:331–377, 2014.

Joseph Farrell and Paul Klemperer. Coordination and lock-in: Competition with switching costs and network effects. In Mark Armstrong and R. Porter, editors, *Handbook of Industrial Organization*, volume 3, pages 1967–2072. Elsevier B.V., 2007.

Ronald E. Frank. Brand choice as a probability process. *The Journal of Business*, 35(1):43–56, 1962.

Shane Frederick, George Loewenstein, and Ted O'Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401, 2002.

Brett R Gordon and Baohong Sun. A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science*, 34(3):452–470, 2015.

Peter M. Guadagni and John D.C. Little. A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2:203–238, 1983.

Benjamin Handel. Adverse selection and switching costs in health insurance markets: When nudging hurts. *American Economic R*, 103:2643–2682, 2013.

James J. Heckman. *Structural Analysis of Discrete Data and Econometric Applications*, chapter Statistical Models for Discrete Panel Data. The MIT Press, 1981.

James J. Heckman and B. Singer. The identifiability of the proportional hazard model. *Review of Economic Studies*, 51:231–241, 1984.

James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, 6: 4779–4874, 2007.

James J Heckman, Sergio Urzua, and Edward Vytlacil. Instrumental variables in models with multiple outcomes: The general unordered case. *Annales d'Economie et de Statistique*, pages 151–174, 2008.

Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006.

Bo Honore and Ekaterini Kyriazidou. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68:839–874, 2000.

Bo E Honoré and Elie Tamer. Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629, 2006.

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

Mariano Irace. Patient loyalty in hospital choice: Evidence from new york. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2018-52), 2018.

Michael P. Keane. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics*, 15(3):310–327, 1997.

Paul Klemperer. Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies*, 62:515–539, 1995.

Alfred Arthur Kuehn. *An analysis of the dynamics of consumer behavior and its implications for marketing management*. PhD thesis, Carnegie Institute of Technology, 1958.

Sokbae Lee and Bernard Salanié. Identifying effects of multivalued treatments. *Econometrica*, 86 (6):1939–1963, 2018.

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

William F. Massy. Order and homogeneity of family specific brand-switching processes. *Journal of Marketing Research*, 3(1):48–54, 1966.

Whitney K. Newey and Daniel McFadden. Large sample estimation and hypothesis testing. In Robert F. Engle and Daniel L. McFadden, editors, *Handbook of Econometrics*, volume IV, pages 2111–2245. Elsevier Science B.V., 1994.

Ariel Pakes, Jack R Porter, Mark Shepard, and Sophie Calder-Wang. Unobserved heterogeneity, state dependence, and health plan choices. Technical report, National Bureau of Economic Research, 2021.

Maria Polyakova. Regulation of insurance with adverse selection and switching costs: Evidence from medicare part d. *American Economic Journal: Applied Economics*, 8(3):165–95, 2016.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987.

John Rust. Structural estimation of markov decision processes. In Robert F. Engle and Daniel L. McFadden, editors, *Handbook of Econometrics*, volume IV, pages 3081–3143. Amsterdam, London and New York: Elsevier, North-Holland, 1994.

P.B. Seetharaman, Andrew Ainslie, and Pradeep Chintagunta. Investigating household state dependence effects across categories. *Marketing Science*, 36:488–500, 1999.

Andrey Simonov, Jean-Pierre Dubé, Günter Hitsch, and Peter Rossi. State-dependent demand estimation with initial conditions correction. *Journal of Marketing Research*, 57(5):789–809, 2020.

Che-Lin Su and Kenneth L. Judd. Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5):2213–2230, 2012.

Alexander Torgovitsky. Nonparametric inference on state dependence in unemployment. *Econometrica*, 87(5):1475–1505, 2019.

A Appendix

A.1 Proofs of the DPO Model

Proof of Proposition 1: Denote shorthand $U_{it}(-j) \equiv (U_{it}(1),...,U_{it}(j-1),U_{it}(j+1),...,U_{it}(J))$, $U_{it}(-jk) \equiv (U_{it}(1),...,U_{it}(j-1),U_{it}(j+1),...,,U_{it}(k-1),U_{it}(k+1),...,U_{it}(J))$ and $\mathscr{A}^k \equiv \mathscr{J}^k \setminus \{\underbrace{1,...,1}_{k}\}$ for $k \in 1,2,...,J$.Note that if $P \in \mathscr{P}^\star$, then

$$\begin{aligned}
HLB_{jt} \equiv & \mathbb{P}_P[\exists k : U_{it}(j) = j, U_{it}(k) \neq j] \\
= & \mathbb{P}_P[U_{it}(j) = j, U_{it}(-j) \in \mathscr{A}^{J-1}] \\
= & \mathbb{P}_P[Y_{i(t-1)} = j, U_{it}(j) = j, U_{it}(-j) \in \mathscr{A}^{J-1}] + \mathbb{P}_P[Y_{i(t-1)} \neq j, U_{it}(j) = j, U_{it}(-j) \in \mathscr{A}^{J-1}] \\
= & \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(-j) \in \mathscr{A}^{J-1}] + \mathbb{P}_P[Y_{i(t-1)} \neq j, U_{it}(j) = j, U_{it}(-j) \in \mathscr{A}^{J-1}]
\end{aligned} \tag{A.1}$$

where the first row is the definition of HBL_{jt}, the first equality is re-writing the HBL pattern using the shorthand, the third equality follows because under the DPO model, $[Y_{i(t-1)} = j, U_{it}(j) = j]$ if and only if $[Y_{i(t-1)} = j, Y_{it} = j]$.

When $|\mathscr{J}| \geq 3$, we can decompose HBL_{jt} to be

$$\begin{aligned}
HBL_{jt} = & \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(-j) \in \mathscr{A}^{J-1}] + \sum_{l \neq j} \mathbb{P}_P[Y_{i(t-1)} = l, U_{it}(j) = j, U_{it}(-j) \in \mathscr{A}^{J-1}] \\
= & \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(-j) \in \mathscr{A}^{J-1}] \\
& + \sum_{l \neq j} \mathbb{P}_P[Y_{i(t-1)} = l, U_{it}(l) = j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{A}^{J-2}] \\
& + \sum_{l \neq j} \mathbb{P}_P[Y_{i(t-1)} = l, U_{it}(l) \neq j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{J}^{J-2}] \\
= & \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(-j) \in \mathscr{A}^{J-1}] \\
& + \sum_{l \neq j} \mathbb{P}_P[Y_{i(t-1)} = l, Y_{it} = j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{A}^{J-2}] \\
& + \sum_{l \neq j} \mathbb{P}_P[Y_{i(t-1)} = l, Y_{it} \neq j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{J}^{J-2}]
\end{aligned}$$

where the second equality comes from re-writing the summation term using shorthand defined above, and the third equality follows because $[Y_{i(t-1)} = l, U_{it}(l) = j]$ if and only if $[Y_{i(t-1)} = l, Y_{it} = j]$, and $[Y_{i(t-1)} = l, U_{it}(l) \neq j]$ if and only if $[Y_{i(t-1)} = l, Y_{it} \neq j]$. Since the only restrictions put on the three terms are the following for all $l \neq j$,

$$
\begin{aligned}
&0 \leq \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(k) = k] \leq \mathbb{P}[Y_{i(t-1)} = j, Y_{it} = j] \\
&0 \leq \mathbb{P}_P[Y_{i(t-1)} = l, Y_{it} = j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{A}^{J-2}] \leq \mathbb{P}[Y_{i(t-1)} = l, Y_{it} = j] \\
&0 \leq \mathbb{P}_P[Y_{i(t-1)} = l, Y_{it} \neq j, U_{it}(j) = j, U_{it}(-jl) \in \mathscr{J}^{J-2}] \leq \mathbb{P}[Y_{i(t-1)} = l, Y_{it} \neq j]
\end{aligned}
$$

and there is no restriction put across the terms, we can find a $P \in \mathscr{P}^\star$ that attains all upper bounds and one that attains all lower bounds. Therefore, $\Theta^\star = [0, \mathbb{P}[Y_{i(t-1)} = j, Y_{it} \neq j] + \mathbb{P}[Y_{i(t-1)} \neq j]]$.

When $|\mathscr{J}| = 2$, the same logic follows with a bit modification. Let $k = \mathscr{J} \setminus j$ and we can simplify HBL_{jt} to be

$$
\begin{aligned}
HBL_{jt} =& \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(k) = k] + \mathbb{P}_P[Y_{i(t-1)} = k, U_{it}(j) = j, U_{it}(k) = k] \\
=& \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(k) = k] + \mathbb{P}_P[Y_{i(t-1)} = k, Y_{it} = k, U_{it}(j) = j] \qquad \text{(A.2)}
\end{aligned}
$$

Since observational equivalence requires that

$$
\begin{aligned}
&0 \leq \mathbb{P}_P[Y_{i(t-1)} = j, Y_{it} = j, U_{it}(k) = k] \leq \mathbb{P}[Y_{i(t-1)} = j, Y_{it} = j] \\
&0 \leq \mathbb{P}_P[Y_{i(t-1)} = k, Y_{it} = k, U_{it}(j) = j] \leq \mathbb{P}[Y_{i(t-1)} = k, Y_{it} = k] \qquad \text{(A.3)}
\end{aligned}
$$

and there is no restriction put between the two terms, $\Theta^\star = [0, \mathbb{P}[Y_{i(t-1)} = j, Y_{it} = j] + \mathbb{P}[Y_{i(t-1)} = k, Y_{it} = k]]$ when $|\mathscr{J}| = 2$. □

Proof of Proposition 2: The stationarity of potential outcomes implied by the RUM depends on the stationarity of counterfactual state variables (p_t, ε_t). When price covariates and random utility shocks are stationary, potential outcomes have a stationarity distribution. For every $t, t' \geq 1$, every

$u \equiv (u_1, ..., u_J) \in \{1, ..., J\}^J$,

$$
\begin{aligned}
&\mathbb{P}_P[U_{it}(1) = u_1, ..., U_{it}(J) = u_J] \\
&= \mathbb{P}\left[\arg\max_k \{v_k(p_t, 1) + \varepsilon_{kt}\} = u_1, ..., \arg\max_k \{v_k(p_t, J) + \varepsilon_{kt}\} = u_J\right] \\
&= \mathbb{E}\left[\prod_{k=1}^{J} \mathbb{1}[u_k = \arg\max_j \{v_j(p_t, k) + \varepsilon_{jt}\}]\right] \\
&= \int \prod_{k=1}^{J} \mathbb{1}[u_k = \arg\max_j \{v_j(p_t, k) + \varepsilon_{jt}\}] d\mathbf{F}_{p_t} d\mathbf{F}_{\varepsilon_t} \\
&= \int \prod_{k=1}^{J} \mathbb{1}[u_k = \arg\max_j \{v_j(p_{t'}, k) + \varepsilon_{jt'}\}] d\mathbf{F}_{p_t'} d\mathbf{F}_{\varepsilon_t'} \\
&= \mathbb{P}\left[\arg\max_k \{v_k(p_{t'}, 1) + \varepsilon_{kt'}\} = u_1, ..., \arg\max_k \{v_k(p_{t'}, J) + \varepsilon_{kt'}\} = u_J\right] \\
&= \mathbb{P}_P[U_{it'}(1) = u_1, ..., U_{it'}(J) = u_J] \qquad \text{(A.4)}
\end{aligned}
$$

□

Proof of Proposition 3: A sufficient condition for Assumption MTS is A sufficient condition for Assumption MTS is

$$
\mathbb{P}_P[U_{it}(y) = j, Y_{i(t-1)} = j | Y_{i(t-2)} = \tilde{y}] \geq \mathbb{P}_P[U_{it}(y) = j | Y_{i(t-2)=\tilde{y}}] \cdot \mathbb{P}[U_{i(t-1)} = j | Y_{i(t-2)} = \tilde{y}] \qquad \text{(A.5)}
$$

This is because if A and B are two events and B occurs with probability strictly within $(0, 1)$, then $\mathbb{P}(A, B) \geq \mathbb{P}(A)\mathbb{P}(B) \Rightarrow \mathbb{P}[A|B] \geq \mathbb{P}[A|B^{\complement}]$.

Define shorthand $g(\varepsilon) \equiv \varepsilon_j - \varepsilon_k$, $\Omega(p, y) \equiv v_k(p, y) - v_j(p, y) = \gamma_k - \gamma_j + \alpha(p_k - p_j) + \lambda(\mathbb{I}_{[l=k]} - \mathbb{I}_{[l=j]})$. Then Assumption MTS is satisfied if

$$
\begin{aligned}
&\mathbb{P}[g(\varepsilon_t) \geq \Omega(p, y), g(\varepsilon_{t-1}) \geq \Omega(p, \tilde{y}) | Y_{i(t-2)} = \tilde{y}] \qquad \text{(A.6)} \\
&\quad \geq \mathbb{P}[g(\varepsilon_t) \geq \Omega(p, y) | Y_{i(t-2)} = \tilde{y}] \cdot \mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p, \tilde{y}) | Y_{i(t-2)} = \tilde{y}] \qquad \text{(A.7)}
\end{aligned}
$$

Since $\varepsilon_t \perp p$ and $(g(\varepsilon_t), g(\varepsilon_{t-1}))$ are positvely correlated,

$$\begin{aligned}
&\mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y), g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|Y_{i(t-2)} = \tilde{y}] \\
&\quad = \mathbb{E}\left[\mathbb{P}\left[g(\varepsilon_t) \geq \Omega(p,y), g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|Y_{i0}, \varepsilon_{1:(t-2)}, p, Y_{i(t-2)}\right]|Y_{i(t-2)} = \tilde{y}\right] \\
&\quad = \mathbb{E}\left[\mathbb{P}\left[g(\varepsilon_t) \geq \Omega(p,y), g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|Y_{i0}, \varepsilon_{1:(t-2)}\right]|Y_{i(t-2)} = \tilde{y}\right] \\
&\quad \geq \mathbb{E}\left[\mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y)|Y_{i0}, \varepsilon_{1:(t-2)}] \cdot \mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|Y_{i0}, \varepsilon_{1:(t-2)}]|Y_{i(t-2)} = y\right] \qquad \text{(A.8)}
\end{aligned}$$

where the first equality comes from law of iterated expectation, the second equality comes from (1) $Y_{i(t-2)}$ being completely determined by ε, p, and Y_{i0} and (2) $\varepsilon_t \perp p$, and the first inequality comes from $g(\varepsilon_t), g(\varepsilon_{t-1})$ being positively correlated.

Since $\Omega(p,y)$ is decreasing in $p_k - p_j$ for all y, we have both $\mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y)|\varepsilon_{1:(t-2)}]$ and $\mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|\varepsilon_{1:(t-2)}]$ are increasing in $p_k - p_j$, so that these two terms have positive covariance (Lehmann 1966). Building on (A.8),

$$\begin{aligned}
&\mathbb{P}[\varepsilon_{jt} - \varepsilon_{kt} \geq v_k(p,y) - v_j(p,y), \varepsilon_{j(t-1)} - \varepsilon_{k(t-1)} \geq v_k(p,\tilde{y}) - v_j(p,\tilde{y})|Y_{i(t-2)} = \tilde{y}] \\
&\quad \geq \mathbb{E}\left[\mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y)|\varepsilon_{1:(t-2)}] \cdot \mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|\varepsilon_{1:(t-2)}]|Y_{i(t-2)} = y\right] \\
&\quad \geq \mathbb{E}\left[\mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y)|\varepsilon_{1:(t-2)}]|Y_{i(t-2)} = \tilde{y}\right] \cdot \mathbb{E}\left[\mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|\varepsilon_{1:(t-2)}]|Y_{i(t-2)} = \tilde{y}\right] \\
&\quad = \mathbb{P}[g(\varepsilon_t) \geq \Omega(p,y)|Y_{i(t-2)} = \tilde{y}] \cdot \mathbb{P}[g(\varepsilon_{t-1}) \geq \Omega(p,\tilde{y})|Y_{i(t-2)} = \tilde{y}] \qquad \text{(A.9)}
\end{aligned}$$

□

Proof of Proposition 4: A sufficient condition for Assumption MTS is

$$\mathbb{P}_P[U_{it}(y) = j, Y_{i(t-1)} = j|Y_{i(t-2)} = \tilde{y}] \geq \mathbb{P}_P[U_{it}(y) = j|Y_{i(t-2)=\tilde{y}}] \cdot \mathbb{P}[U_{i(t-1)} = j|Y_{i(t-2)} = \tilde{y}] \qquad \text{(A.10)}$$

This is because if A and B are two events and B occurs with probability strictly within $(0,1)$, then

$$\begin{aligned}
\mathbb{P}(A,B) \geq& \mathbb{P}(A)\mathbb{P}(B) \\
\Longrightarrow \mathbb{P}[A|B] = \frac{\mathbb{P}[A,B]}{\mathbb{P}[B]} \geq& \mathbb{P}[A] \geq \frac{\mathbb{P}[A] - \mathbb{P}[A,B]}{1 - \mathbb{P}[B]} = \mathbb{P}[A|B^{\complement}] \qquad \text{(A.11)}
\end{aligned}$$

Therefore, Assumption MTS is satisfied if

$$\begin{aligned}&\mathbb{P}[\arg\max_k\{v_k(p_t,y)+\varepsilon_{kt}\}=j,\arg\max_k\{v_k(p_{t-1},\tilde{y})+\varepsilon_{k(t-1)}\}=j|Y_{i(t-2)}=\tilde{y}]\\&\quad\geq\mathbb{P}[\arg\max_k\{v_k(p_t,y)+\varepsilon_{kt}\}=j|Y_{i(t-2)}=\tilde{y}]\cdot\mathbb{P}[\arg\max_k\{v_k(p_{t-1},\tilde{y})+\varepsilon_{k(t-1)}\}=j|Y_{i(t-2)}=\tilde{y}]\end{aligned} \tag{A.12}$$

Define shorthand $\Omega(\varepsilon,p,y)\equiv\varepsilon_{jt}-\max_k\{v_k(p,y)-v_j(p,y)+\varepsilon_{kt}\}$. Then Assumption MTS is satisfied if

$$\begin{aligned}&\mathbb{P}[\Omega(\varepsilon_t,p_t,y)\geq 0,\Omega(\varepsilon_{t-1},p_{t-1},\tilde{y})\geq 0|Y_{i(t-2)}=\tilde{y}]\\&\quad\geq\mathbb{P}[\Omega(\varepsilon_t,p_t,y)\geq 0|Y_{i(t-2)}=\tilde{y}]\cdot\mathbb{P}[\Omega(\varepsilon_{t-1},p_{t-1},\tilde{y})\geq 0|Y_{i(t-2)}=\tilde{y}]\end{aligned} \tag{A.13}$$

Since $\varepsilon_t\sim i.i.d.\ F_\varepsilon(\varepsilon)$ and prices are time-invariant such that $p_t=p_{t-1}=p$

$$\begin{aligned}&\mathbb{P}[\Omega(\varepsilon_t,p,y)\geq 0,\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0|Y_{i(t-2)}=\tilde{y}]\\&\quad=\mathbb{E}\left[\mathbb{1}\left[\Omega(\varepsilon_t,p,y)\geq 0,\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0\right]|Y_{i(t-2)}=\tilde{y}\right]\\&\quad=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left[\Omega(\varepsilon_t,p,y)\geq 0,\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0\right]|Y_{i(t-2)}=\tilde{y},Y_{i0},\varepsilon_{1:(t-2)},p\right]|Y_{i(t-2)}=\tilde{y}\right]\\&\quad=\mathbb{E}\left[\mathbb{P}\left[\Omega(\varepsilon_t,p,y)\geq 0,\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0|Y_{i0},\varepsilon_{1:(t-2)},p\right]|Y_{i(t-2)}=\tilde{y}\right]\\&\quad=\mathbb{E}\left[\mathbb{P}\left[\Omega(\varepsilon_t,p,y)\geq 0\right]\cdot\mathbb{P}\left[\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0\right]|Y_{i(t-2)}=\tilde{y}\right]\end{aligned} \tag{A.14}$$

where the second equality comes from the law iterated expectation, the third equality comes from $Y_{i(t-1)}$ being fully determined by the initial condition Y_{i0}, the history of ε, and p, the last equality comes from ε_t being $i.i.d.$

Note that $\mathbb{P}[\Omega(\varepsilon_t,p,y)\geq 0]$ is the utility -maximizing choice probability $\sigma_j(p,y)$ and similarly $\mathbb{P}[\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0]=\sigma_j(\tilde{y},p)$, then (A.14) can be written as

$$\begin{aligned}&\mathbb{P}[\Omega(\varepsilon_t,p,y)\geq 0,\Omega(\varepsilon_{t-1},p,\tilde{y})\geq 0|Y_{i(t-2)}=\tilde{y}]\\&\quad=\mathbb{E}\left[\sigma_j(p,y)\cdot\sigma_j(p,\tilde{y})|Y_{i(t-2)}=\tilde{y}\right]\end{aligned} \tag{A.15}$$

If choice probabilities $\sigma_j(y,p)$ and $\sigma_j(p,\tilde{y})$ have weakly positive covariance, we can build on

(A.14)

$$
\begin{aligned}
&\mathbb{P}[\arg\max_k\{v_k(p,y)+\varepsilon_{kt}\}=j, \arg\max_k\{v_k(p,\tilde{y})+\varepsilon_{k(t-1)}\}=j|Y_{i(t-2)}=\tilde{y}] \\
&\geq \mathbb{E}\left[\sigma_j(p,y)|Y_{i(t-2)}=\tilde{y}\right]\cdot\mathbb{E}\left[\sigma_j(p,\tilde{y})|Y_{i(t-2)}=\tilde{y}\right] \\
&= \mathbb{P}[\arg\max_k\{v_k(p,y)+\varepsilon_{kt}\}=j|Y_{i(t-2)}]\cdot\mathbb{P}[\arg\max_k\{v_k(p,\tilde{y})+\varepsilon_{k(t-1)}\}=j|Y_{i(t-2)}=\tilde{y}]
\end{aligned}
\tag{A.16}
$$

where the inequality comes from (A.14), (A.15) and the weakly positive covariance of choice probabilities, and the equality comes frome reversing back the construction of $\Omega(\cdot)$. Therefore, the sufficient condition (A.12) for Assumption MTS is satisfied. □

Proof of Proposition 5: A sufficient condition for Assumption MTR is that

$$
\mathbb{P}\left[\mathbb{1}[j=\arg\max_l\{v_l(p_t,j)+\varepsilon_{lt}\}]\geq\mathbb{1}[j=\arg\max_l\{v_l(p_t,k)+\varepsilon_{lt}\}]\right]=1 \tag{A.17}
$$

Note that in the RUM model, for $j,k,l\in\mathscr{D}_{-0}$

$$
\begin{aligned}
v_l(p,j)-v_l(p,k) &= u_l(p,j)+\beta\mathbf{F}_l(p,j)(\mathbf{m}+\mathbf{v}_0)-u_l(p,k)-\beta\mathbf{F}_l(p,k)(\mathbf{m}+\mathbf{v}_0) \\
&= u_l(p,j)-u_l(p,k)+\beta(\mathbf{F}_l(p,j)-\mathbf{F}_l(p,k))(\mathbf{m}+\mathbf{v_0}) \\
&= u_l(p,j)-u_l(p,k)
\end{aligned}
\tag{A.18}
$$

where the last equality comes from $\mathbf{F}_l(p,j)-\mathbf{F}_l(p,k)=0$ under the assumption that the price evolution is exogeneous, and literature convention that loyalty evolves deterministically after a purchase.

If $u_j(p,j)-u_j(p,k)\geq 0$ and $u_l(p,j)-u_l(p,k)\leq 0$ at every p for all $k,l\in\mathscr{D}\setminus\{0,j\}$, then exogenously switching loyalty state from k to j weakly increases the choice-specific value function of j and weakly decreases choice-specific value functions of all other alternatives. Therefore, $\mathbb{1}[j=\arg\max_l\{v_l(p_t,j)+\varepsilon_{lt}\}]\geq\mathbb{1}[j=\arg\max_l\{v_l(p_t,k)+\varepsilon_{lt}\}]$ must hold. □

A.2 DDC and the exclusion restriction

A consumer has perfect substitutes preferences over a set of $J = 1,...,J$ market goods in a commodity group and a $j = 0$ outside good:

$$U(\mathbf{q},z) = \tilde{U}(\sum_{j\in\mathscr{D}} \gamma_j q_j, z)$$

where z is an essential numeraire[10] with $p_z = 1$, q_j is the quantity consumed of product j, ψ_j is the constant marginal utility per unit of product j, α is the marginal utility for the numeraire (i.e., all other consumption expenditure), and y is her budget. Assume $\gamma_0 = 0$ and $p_0 = 0$. The perfect substitutes preferences is necessary for the consumer to select at most one of the product alternatives (discrete choice). If the consumer also has unit demand for the products, $q_j \in 0,1$, $\forall j \in \mathscr{D}$, then she will purchase one unit of the product with the highest of the following choice-specific values

$$v_j = \tilde{U}(\gamma_j, y - p_j),\ \forall j \in \mathscr{D}.$$

These choice-specific values give rise to the standard discrete-choice model of demand. As can be seen, the choice-specific value of a product j does not depend on any of hte other prices, $\mathbf{p_{-j}}$.

A.3 MPEC Estimator for DDC

In the empirical data, we observe $i = 1,...,N$ choice occasions during which a consumer makes a discrete choice $j = in\mathscr{D}$. Define the vector of outcomes, $Y_i = (Y_{i0},...,Y_{iJ})$ where $Y_{ij} \in 0,1$ indicates whether the consumer chooses j. From the empirical DDC model in section 5.3, a consumer's choice-specific values are:

$$v_j(s;\theta,\beta) = \gamma_j - \alpha p_j + \lambda \mathbb{I}_{\{\ell=j\}} + \beta\pi\mathbf{F}_j(s)\mathbf{B}_j(\beta)\ln\left\{\sum_{k\in\mathscr{D}}\exp(\mathbf{v}_k)\right\}$$

where $\theta = (\gamma_1,...,\gamma_J,\alpha,\lambda)$ is the vector of taste parameters, $\beta \in [0,1)$ is the discount factor, $\mathbf{F}_j$is the $L\times L$ state transition matrix conditional on current choice j and $\mathbf{B}_j(\beta) = \left[\mathbf{I} - \beta(1-\pi)\mathbf{F}_j\right]^{-1}$

[10]We assume (.,.) is quasi-linear in its second argument to ensure the consumer always purchases a strictly positive quantity of the numeraire good, z.

is an $(L\times L)$ matrix. Let $\Gamma=(\beta,\theta)$ denote all the structural parameters.

Let $\mathbf{v}=\left(v_{j1},v_{j2},....,v_{J+1,L}\right)$ denote the choice-specific values in each of the L states. We can now define an MPEC estimator:

$$\Theta^{*}=\underset{\Theta=(\Gamma,\mathbf{v})}{\arg\max}\sum_{i=1}^{N}\sum_{j=0}^{J}\left(Y_{ij}\ln\left(\sigma_{j}\left(\mathbf{p}_{i},\ell_{i};\theta,\mathbf{v}\right)\right)\right)$$

subject to the constraints

$$G_{js}(\Theta)=v_{js}-u_{j}(s;\Gamma)-\beta\pi\mathbf{F}_{j}(s)\mathbf{B}_{j}(\beta)\ln\left\{\sum_{k\in\mathscr{D}}\exp\left(\mathbf{v}_{k}\right)\right\}\ ,\forall j\in\mathscr{D},\forall s\in\mathscr{S}$$

where

$$\sigma_{j}(\mathbf{p}_{i},\ell_{i};\theta,\mathbf{v})=\frac{\exp\left(u_{j}\left(\mathbf{p}_{i},\ell_{i};\theta\right)+v_{js_{i}}-u_{j}\left(s_{i};\theta\right)\right)}{\sum_{k\in\mathscr{D}}\exp\left(u_{k}\left(\mathbf{p}_{i},\ell_{i};\theta\right)+v_{ks_{i}}-u_{k}\left(s_{i};\theta\right)\right)}$$

and

$$u_{j}(\mathbf{p},\ell;\theta)=\gamma_{j}-\alpha p_{j}+\lambda\mathbb{I}_{\{\ell=j\}}.$$

Note that we only discretize the expected future prices. So we evaluate the flow utility at the observed prices when we write $u_{j}(\mathbf{p}_{i},\ell;\theta)$ and we use the cluster classification for a given price when we write $u_{j}(s_{i};\theta)$.

We also provide gradients to improve the robustness and speed of the outer-loop optimization in the MPEC estimator. For the objective function, we use automatic differentiation.[11]. To improve speed, we supply code for the constraint Jacobian. A useful result for the derivation of the constraint Jacobian is that $\nabla_{\beta}\mathbf{B}_{j}\left(\beta\right)=\mathbf{B}_{j}\left(\beta\right)\left(1-\pi\right)F_{j}\left(s\right)\mathbf{B}_{j}\left(\beta\right)$. The Jacobian elements are:

$$\begin{aligned}
\frac{\partial G_{js}(\Theta)}{\partial\beta} &= -\pi\mathbf{F}_{j}(s)\mathbf{B}_{j}(\beta)\ln\left\{\sum_{k\in\mathscr{D}}\exp\left(\mathbf{v}_{k}\right)\right\}-\beta\pi\mathbf{F}_{j}(s)\mathbf{B}_{j}(\beta)(1-\pi)\mathbf{F}_{j}(s)\mathbf{B}_{j}(\beta\ln\left\{\sum_{k\in\mathscr{D}}\exp\left(\mathbf{v}_{k}\right)\right\}\\
\frac{\partial G_{js}(\Theta)}{\partial\theta_{k}} &= -x_{sjk}\\
\frac{\partial G_{js}(\Theta)}{\partial v_{js}} &= 1-\beta\pi\mathbf{F}_{js}(s)\mathbf{B}_{j}(\beta)\sigma_{js},\forall j\in\mathscr{D},\forall s\in\mathscr{S}\\
\frac{\partial G_{js}(\Theta)}{\partial v_{ns'}} &= -\beta\pi\mathbf{F}_{js'}(s)\mathbf{B}_{j}(\beta)\sigma_{ns'},\forall j\in\mathscr{D},\forall n\in\mathscr{D}_{-j},\forall s,s'\in\mathscr{S}
\end{aligned}$$

where $\sigma_{js}\equiv\sigma_{j}(s;\theta,\mathbf{v})$

[11] We use the *dl* package in matlab to obtain exact derivatives

A.4 Group Fixed-Effects and Unobserved Heterogeneity

To control for the well-known persistent, unobserved heterogeneity in consumer tastes for CPG products (e.g., Allenby and Rossi, 1998). We implement a *grouped fixed effects* (GFE) approach (Bonhomme et al., 2021). This approach allows us to estimate the mean effects of our structural parameters,Θ, in a way that is robust to unspecicied between-consumer heterogeneity.

Let consumers be indexed by $h = 1, ..., H$. We use a k-means clustering algorithmWe use Matlab's *kmeans* function to classify each consumer into groups based on the vector of data moments, w_h, that are informative about the heterogeneity. Given our moderate-sized cross-section, we use $K = 3$ groups. The moments consider of hosuehold's demographics and shopping summaries. The demographics consist of income, household size, age of primary shopper, highschool completion dummy, college completion dummy, retired dummy, unemployed dummy, presence of dogs, presence of cats, presence of children, marital status, hispanic dummy, white dummy and black dummy. The shopping moments consist of each household's propensity to purchase each of the available brands.

Standard errors are calculated by block-bootstrapping consumers and re-running the kmean clustering.

A.5 Computing the HBL statistic under the DDC

To conform with the HBL measurement under the DPO, We use only those trips during which a purchase occurs. To simulate choices in each observed state, we first use the DDC estimates to compute the predicted choice-specific values at each observed state, $\mathbf{s}_{it}$ where $i = 1, ..., N$ and $t = 1, ..., T$: $\{\hat{v}_{ilt}(\mathbf{s}_{it})\}_{l\in\mathscr{D}}$. We then simulate random utilities ε_{ilt} for each choice alternative $l \in \mathscr{J}$, consumer i and time period t. We then approximate HBL_j as follows:

$$HBL_j = \mathbb{P}_\varepsilon[\exists k \neq j \in \mathscr{J} \text{ s.t. } j = \underset{l\in\mathscr{J}}{\operatorname{argmax}}\{v_{il}(\mathbf{p}_t, j) + \varepsilon_{ilt}\}, \text{ and } j \neq \underset{l\in\mathscr{J}}{\operatorname{argmax}}\{v_{il}(\mathbf{p}_t, k) + \varepsilon_{ilt}\}, \forall i, t] \tag{A.19}$$

$$\approx \frac{1}{NTR}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{R}\mathbb{1}[j = \underset{l\in\mathscr{J}}{\operatorname{argmax}}\{v_{il}(\mathbf{p}_t, j) + \varepsilon^r_{ilt}\}, \text{ and } j \neq \underset{l\in\mathscr{J}}{\operatorname{argmax}}\{v_{il}(\mathbf{p}_t, k) + \varepsilon^r_{ilt}\}, \forall i, t].$$ 2005/06/28v

To implement HBL, we use $R = 10,000$ simulation draws. Intuitively, HBL_j measures the fraction of trips (conditional on purchase) for which a consumer would repeat-buy brand j when loyal to it; but would buy a different brand under at least one alternative loyalty state.

A.6 Computing the Long-Run Elasticity of Demand

To compute the long-run elasticity of demand, we compute the expected change in the steady-state choice probabilities for each product when we permanently increase the price of the top-selling brand, k. Formally, we compute the factual expected choice probabilities in equilibrium, $\Pi\sigma(\mathbf{p},\ell)$, where $\sigma(\mathbf{p},\ell)$ is the $(L \times J+1)$ matrix of choice probabilities in each of the L states and Π is the $(1 \times L)$ row-vector of equilibrium probabilities of being in each state. We then create a counterfactual set of L states which are identical to $(\mathbf{p},\ell)$ except for the modification for product k: $\tilde{p}_{kr} = 1.1 p_{kr}$. We then compute the corresponding counter-factual expected choice probabilities in equilibrium, $\tilde{\Pi}\sigma(\tilde{\mathbf{p}},\ell)$ where $\tilde{\Pi}$ is the vector of counter-factual equilibrium probabilities of being in each state. The long-run elasticity of demand is then:

$$\varepsilon_j^{LR} = \frac{\frac{\sigma(\tilde{\mathbf{p}},\ell)\tilde{\Pi} - \sigma(\mathbf{p},\ell)\Pi}{\sigma(\mathbf{p},\ell)\Pi}}{0.1}. \tag{A.21}$$

We compute the steady-state probabilities as follows. Let $\mathbf{Q}$ be the $(R \times J)$ transition matrix for prices. We compute the $(1 \times J)$ row=vector of steady-state price probabilities, $\mathbf{q} = \mathbf{q}\mathbf{Q}$, by taking the left eigenvector of $\mathbf{Q}$ corresponding to the eigenvalue 1. We then use forward-simulation to find the steady-state probabilities Π corresponding to the combined state, $(\mathbf{p},\ell)$,with transition matrix, $\mathbf{F}$. We repeat the following steps $r = 1,...,R$ times:

1. Initialize prices by assuming they are in equilibrium and drawing $\mathbf{p}_0^r \in \mathscr{S}$ using the probabilities $\mathbf{q}$. Then simulate a sequence of prices of length T, $\mathbf{p}^r \equiv (\mathbf{p}_1^r,...,\mathbf{p}_T^r)$, using the transition process $\mathbf{Q}$.

2. Initialize the loyalty state by drawing $\ell_0 \in \mathscr{J}$ from uniform discrete distribution. Then use the sequence of prices, $\mathbf{p}^r$, and the DDC to simulate a choice sequence of length T and the corresponding sequence of loyalty states, $\ell^r = (\ell_1^r,...,\ell_T^r)$.

We can then approximate the steady-state probabilities, Π, as follows:

$$\Pi_l = \frac{1}{R} \sum_{r=1}^{R} \mathbb{1}\left[s_T^r = s_l\right], \ \forall l \in \mathscr{S}. \tag{A.22}$$

We use $R = 3000$ and $T = 200$. Results for the CPG categories are displayed in tables A1 to A8.

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
FOLGERS_S	-3.257	-3.391	4.1
FOLGERS_L	0.13	0.214	65
MAXWELL_S	0.142	0.234	64.5
MAXWELL_L	0.146	0.235	60.4
PRIVATELABEL_S	0.145	0.236	62.4
no purch	0.164	0.16	-2

Table A1: LR elasticities: Coffee

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
DEGREE	-1.277	-1.621	27
DOVE	0.009	0.198	2189.1
MENNEN	0.009	0.174	1844.3
OLDSPICE	0.009	0.178	1847.8
RIGHT	0.009	0.17	1750.2
SECRET	0.009	0.17	1707.5
no purch	0.013	0.014	14.5

Table A2: LR elasticities: Deodorant

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
HELLMANNS	-2.581	-3.011	16.7
PRIVATELABEL	0.045	2.115	4612.2
KRAFTMAYO	0.048	1.811	3637
no purch	0.088	0.1	14.6

Table A3: LR elasticities: Mayonnaise

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
ICANTBELIEVEITSNOTBUTTER	-1.228	-1.253	2
SMARTBALANCE	0.012	0.034	173.5
SHEDDSCOUNTRYCROCK	0.013	0.045	259.6
PROMISE	0.014	0.042	210.3
no purch	0.044	0.049	11.8

Table A4: LR elasticities: Margarine

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
JIF	-1.446	-1.961	35.6
PETERPAN	0.022	0.389	1667.1
PRIVATELABEL	0.023	0.354	1415.1
SKIPPY	0.025	0.286	1022
no purch	0.032	0.032	2

Table A5: LR elasticities: Peanut Butter

A.7 Estimating DPO Bounds from Theoretical Choice Sequence Probabilities

To assess the performance of the DPO model, we estimate DPO bounds from the theoretical probabilities of choice sequences generated by a Randum Utility Model with known parameters. By using the theoretical probabilities of observing a choice sequence, we can derive DPO bounds as if we have infinite data, and hence separating the estimation problem from the inference problem.

Formally, we assume a data generating process as in (2.3) where the flow utility takes the form $u_j(s) = \gamma_j - \alpha p_j + \lambda \mathbb{1}_{\{\ell_t = j\}}$ and the loyalty state evolution follows (5.1). We calculate the expected probability of observing a choice sequence in equilibrium,

$$
\begin{aligned}
& P[\mathbf{Y}_i = (y_0, \ldots, y_T)] \\
& \quad = \mathbf{Q}_s P[\mathbf{Y}_i = (y_0, \ldots, y_T) | \mathbf{s}_0 = s] \\
& \quad = \mathbf{Q}_s P[Y_{i0} = y_0 | \mathbf{s}_0 = s] \prod_{t=1}^{T} P[Y_{it} = y_t | Y_{i(0:(t-1))} = (y_0, .., y_{t-1}), \mathbf{s}_0 = s]
\end{aligned}
$$

where $\mathbf{Q}_s$ is a vector of equilibrium probabilities of being in state s, $P[Y_{i0} = y_0 | \mathbf{s}_0 = s]$ is a state-specific choice probability conditional on being in the initial state $\mathbf{s}_0$, and $P[Y_{it} = y_t | Y_{i(0:(t-1))}, \mathbf{s}_0]$ is a one-period choice probability conditional on a choice history and the initial state. We can further

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
FRANCESCORINALDI	-3.086	-3.302	7
HUNTS	0.086	0.525	510.1
PREGO	0.089	0.562	528.5
RAGU	0.099	0.577	482.2
no purch	0.129	0.138	7.4

Table A6: LR elasticities: Spaghetti Sauce

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
COLGATE	-1.842	-1.925	4.5
ORALB	0.017	0.22	1174.6
no purch	0.022	0.023	3.9

Table A7: LR elasticities: Toothbrushes

simplify the conditional choice probability term

$$\begin{aligned} & P[Y_{it} = y_t | Y_{i(0:(t-1))}, \mathbf{s}_0] \\ & \quad = \mathbf{G}(\mathbf{p}_t | \mathbf{s}_0) P[Y_{it} = y_t | Y_{i(0:(t-1))}, \mathbf{p}_t, \mathbf{s}_0] \\ & \quad = \mathbf{G}(\mathbf{p}_1 | \mathbf{s}_0) (\mathbf{F}_p)^{t-1} P[Y_{it} = y_t | \mathbf{p}_t, \ell_t] \end{aligned}$$

where F_p is the exogenous Markov price transition distribution, $\mathbf{G}(\mathbf{p}_1|\mathbf{s}_0) = \left[F_p', ..., F_p'\right]'_{L\times|\mathscr{P}|}$ is a one-period transition probability from initial state $\mathbf{s}_0$ to the support points of prices in the next period, and $P[Y_{it} = y_t|\mathbf{p}_t, \ell_t]$ is the choice probability conditional on being in state $(\mathbf{p}_t, \ell_t)$, where ℓ_t is determined recursively by the initial loyalty state ℓ_0 and the chocie history through $\ell_t = \mathbb{1}[y_{t-1} = J+1]\ell_{t-1} + 1[y_{t-1} \neq J+1]y_{t-1}$.

With a known set of parameters of the parametric DDC model, we can calculate theoretical probabilities of all permutations of a choice sequence of length T_1. We further count the probabilities of the last T_2 purchases conditional having at least T_2 purchases in the sequence. We feed the resulting theoretical probability of a choice sequence of length T_2 into the DPO model for estimation.

Table A9 and A10 report the DPO bounds estimated from various data generating processes.

brand	SR Elasticity (RUM)	LR Elasticity (DDC)	% difference
YOPLAIT	-1.402	-1.547	10.4
DANNON	0.05	0.338	573.6
COLOMBO	0.057	0.361	532.9
no purch	0.098	0.111	13.1

Table A8: LR elasticities: Yogurt

β	λ	true HBL	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$
0.000	0.000	0.000	0.000	0.000
0.900	0.000	0.000	0.000	0.000
0.000	1.000	0.322	0.046	0.371
0.900	1.000	0.313	0.003	0.375

This table summarizes the HBL bounds estimated from theoretical probabilities of a two-product choice sequences using a binary DPO model. Choices are generated by an RUM with various β and λ. Choice sequence probabilities are calculated with $T_1 = 15$ and $T_2 = 6$.

Table A9: Theoretical bounds: A 2-product D.G.P.

A.8 Other Tables

β	λ	true HBL	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$
0.000	0.000	0.000	0.000	0.032
0.900	0.000	0.000	0.000	0.032
0.000	1.000	0.346	0.019	0.285
0.900	1.000	0.344	0.011	0.284

(a) Binary DPO Bounds

β	λ	true HBL	$\hat{HBL}_{lb}$	$\hat{HBL}_{ub}$
0.000	0.000	0.000	0.001	0.024
0.900	0.000	0.000	0.001	0.024
0.000	1.000	0.346	0.007	0.474
0.900	1.000	0.344	0.005	0.477

(b) Multinomial DPO Bounds

This table summarizes the HBL bounds estimated from theoretical probabilities of a four-product choice sequences. Choices are generated by an RUM with various β and λ. Choice sequence probabilities are calculated with $T_1 = 10$ and $T_2 = 6$, and by collapsing choices in to (a) the most popular brand and a composite brand and (b) top two brands and a composite brand.

Table A10: Theoretical bounds: A 4-product D.G.P.

brand	HBL	$HBL \mid Y_t = j$	$HBL \mid Y_t = Y_{t-1} = j$
$MAXWELL_S$	[0.137,0.233]	[0.182,0.190]	[0.292,0.305]
	[0.039,0.308]	[0.080,0.314]	[0.128,0.499]
$FOLGERS_S$	[0.032,0.272]	[0.051,0.327]	[0.083,0.532]
	[0.000,0.309]	[0.000,0.396]	[0.000,0.645]
$FOLGERS_L$	[0.054,0.282]	[0.148,0.359]	[0.311,0.787]
	[0.035,0.304]	[0.094,0.375]	[0.194,0.832]
$PRIVATELABEL_S$	[0.025,0.344]	[0.125,0.433]	[0.262,0.914]
	[0.024,0.347]	[0.124,0.437]	[0.261,0.927]
$MAXWELL_L$	[0.000,0.367]	[0.000,0.227]	[0.000,0.986]
	[0.000,0.367]	[0.000,0.228]	[0.000,0.986]

(a) Binary DPO Bounds

brand	HBL	$HBL \mid Y_t = j$	$HBL \mid Y_t = Y_{t-1} = j$
$MAXWELL_S$	[0.070,0.395]	[0.096,0.643]	[0.000,0.865]
	[0.029,0.489]	[0.029,0.804]	[0.000,1.000]
$FOLGERS_S$	[0.078,0.440]	[0.120,0.841]	[0.000,0.995]
	[0.000,0.524]	[0.030,1.000]	[0.000,1.000]
$FOLGERS_L$	[0.067,0.239]	[0.161,0.499]	[0.009,0.726]
	[0.024,0.350]	[0.056,1.000]	[0.000,1.000]
$PRIVATELABEL_S$	[0.034,0.247]	[0.177,0.859]	[0.009,0.964]
	[0.000,0.426]	[0.000,1.000]	[0.000,1.000]
$MAXWELL_L$	[0.066,0.089]	[0.445,0.656]	[0.443,0.943]
	[0.020,0.475]	[0.137,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Coffee category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A11: DPO Estimates from the Coffee cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
MENNEN	[0.107,0.307]	[0.190,0.293]	[0.282,0.435]
	[0.000,0.358]	[0.099,0.415]	[0.147,0.615]
OLD SPICE	[0.086,0.347]	[0.198,0.428]	[0.330,0.707]
	[0.086,0.367]	[0.197,0.471]	[0.329,0.777]
DOVE	[0.058,0.173]	[0.175,0.235]	[0.305,0.409]
	[0.000,0.257]	[0.000,0.542]	[0.000,0.937]
RIGHT	[0.058,0.224]	[0.209,0.371]	[0.446,0.784]
	[0.000,0.273]	[0.000,0.437]	[0.000,0.920]
DEGREE	[0.035,0.301]	[0.167,0.386]	[0.351,0.808]
	[0.000,0.329]	[0.000,0.466]	[0.000,0.971]
SECRET	[0.020,0.281]	[0.152,0.397]	[0.321,0.821]
	[0.000,0.281]	[0.000,0.402]	[0.320,0.838]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
MENNEN	[0.071,0.513]	[0.112,0.936]	[0.000,1.000]
	[0.000,0.617]	[0.056,1.000]	[0.000,1.000]
OLD SPICE	[0.049,0.380]	[0.088,0.718]	[0.007,0.941]
	[0.029,0.474]	[0.009,1.000]	[0.000,1.000]
DOVE	[0.039,0.249]	[0.122,0.728]	[0.000,0.943]
	[0.000,0.329]	[0.000,1.000]	[0.000,1.000]
RIGHT	[0.049,0.256]	[0.149,0.801]	[0.000,0.972]
	[0.000,0.320]	[0.000,1.000]	[0.000,1.000]
DEGREE	[0.051,0.319]	[0.208,0.704]	[0.020,0.945]
	[0.000,0.416]	[0.095,1.000]	[0.000,1.000]
SECRET	[0.050,0.212]	[0.326,0.834]	[0.159,0.940]
	[0.016,0.370]	[0.073,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Deodorant category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A12: DPO Estimates from the Deodorant cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
HELLMANNS	[0.065,0.386]	[0.052,0.391]	[0.058,0.431]
	[0.033,0.429]	[0.030,0.434]	[0.033,0.478]
PRIVATE LABEL	[0.040,0.343]	[0.131,0.274]	[0.207,0.431]
	[0.000,0.401]	[0.000,0.328]	[0.000,0.515]
KRAFT MAYO	[0.033,0.234]	[0.138,0.445]	[0.217,0.705]
	[0.000,0.286]	[0.000,0.639]	[0.000,1.000]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
HELLMANNS	[0.042,0.579]	[0.032,0.649]	[0.001,0.705]
	[0.023,0.671]	[0.018,0.774]	[0.000,0.839]
PRIVATE LABEL	[0.026,0.409]	[0.081,0.713]	[0.003,0.894]
	[0.000,0.494]	[0.000,0.714]	[0.000,1.000]
KRAFT MAYO	[0.028,0.281]	[0.128,1.000]	[0.005,1.000]
	[0.011,0.327]	[0.035,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Mayonnaise category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A13: DPO Estimates from the Mayonnaise cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
SMART BALANCE	[0.045,0.485]	[0.066,0.426]	[0.075,0.480]
	[0.000,0.560]	[0.021,0.540]	[0.024,0.608]
ICBINB	[0.047,0.342]	[0.085,0.429]	[0.100,0.506]
	[0.000,0.432]	[0.044,0.553]	[0.052,0.652]
SHEDDS COUNTRY CROCK	[0.060,0.273]	[0.165,0.329]	[0.203,0.400]
	[0.041,0.392]	[0.091,0.772]	[0.112,0.938]
PROMISE	[0.025,0.342]	[0.112,0.503]	[0.143,0.639]
	[0.000,0.458]	[0.000,0.816]	[0.000,1.000]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
SMART BALANCE	[0.037,0.574]	[0.044,0.707]	[0.002,0.743]
	[0.000,0.706]	[0.009,0.856]	[0.000,1.000]
ICBINB	[0.048,0.466]	[0.061,0.848]	[0.000,0.947]
	[0.027,0.535]	[0.024,1.000]	[0.000,1.000]
SHEDDS COUNTRY CROCK	[0.039,0.339]	[0.112,0.790]	[0.012,0.907]
	[0.013,0.411]	[0.020,1.000]	[0.000,1.000]
PROMISE	[0.016,0.383]	[0.086,1.000]	[0.000,1.000]
	[0.001,0.471]	[0.000,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Margarine category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A14: DPO Estimates from the Margarine cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
SKIPPY	[0.081,0.261]	[0.141,0.270]	[0.219,0.418]
	[0.064,0.288]	[0.112,0.319]	[0.174,0.493]
JIF	[0.036,0.299]	[0.044,0.341]	[0.064,0.491]
	[0.019,0.326]	[0.027,0.372]	[0.038,0.537]
PRIVATE LABEL	[0.058,0.326]	[0.108,0.271]	[0.171,0.428]
	[0.040,0.354]	[0.070,0.304]	[0.112,0.481]
PETER PAN	[0.023,0.230]	[0.083,0.301]	[0.222,0.849]
	[0.022,0.230]	[0.083,0.301]	[0.221,0.849]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
SKIPPY	[0.056,0.485]	[0.087,0.884]	[0.000,0.998]
	[0.046,0.535]	[0.063,0.928]	[0.000,1.000]
JIF	[0.046,0.432]	[0.068,0.741]	[0.000,0.941]
	[0.008,0.488]	[0.051,0.812]	[0.000,1.000]
PRIVATE LABEL	[0.038,0.430]	[0.080,0.700]	[0.000,0.917]
	[0.026,0.474]	[0.056,0.749]	[0.000,1.000]
PETER PAN	[0.036,0.183]	[0.259,1.000]	[0.013,1.000]
	[0.026,0.223]	[0.152,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Peanut Butter category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A15: DPO Estimates from the Peanut Butter cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
RAGU	[0.034,0.324]	[0.035,0.376]	[0.047,0.501]
	[0.000,0.362]	[0.000,0.405]	[0.000,0.540]
FRANCESCO RINALDI	[0.031,0.247]	[0.081,0.257]	[0.160,0.508]
	[0.000,0.280]	[0.000,0.290]	[0.000,0.572]
PREGO	[0.065,0.253]	[0.146,0.319]	[0.238,0.516]
	[0.038,0.289]	[0.090,0.517]	[0.144,0.832]
HUNTS	[0.030,0.272]	[0.152,0.299]	[0.287,0.569]
	[0.000,0.311]	[0.045,0.476]	[0.000,0.909]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
RAGU	[0.029,0.493]	[0.030,0.702]	[0.000,0.837]
	[0.000,0.536]	[0.000,0.777]	[0.000,0.892]
FRANCESCO RINALDI	[0.031,0.345]	[0.070,0.695]	[0.000,0.929]
	[0.000,0.383]	[0.000,0.809]	[0.000,1.000]
PREGO	[0.046,0.349]	[0.100,0.886]	[0.000,1.000]
	[0.020,0.399]	[0.046,1.000]	[0.000,1.000]
HUNTS	[0.034,0.314]	[0.158,0.832]	[0.000,0.994]
	[0.014,0.360]	[0.054,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Spaghetti Sauce category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A16: DPO Estimates from the Spaghetti Sauce cateogry

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
COLGATE	[0.162,0.242]	[0.085,0.211]	[0.101,0.252]
	[0.000,0.382]	[0.000,0.425]	[0.000,0.506]
ORAL B	[0.162,0.242]	[0.283,0.285]	[0.402,0.405]
	[0.000,0.382]	[0.000,0.590]	[0.000,0.839]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t = j$	$HBL\|Y_t = Y_{t-1} = j$
COLGATE	[0.162,0.242]	[0.085,0.211]	[0.101,0.251]
	[0.053,0.423]	[0.000,0.499]	[0.000,0.593]
ORAL B	[0.162,0.242]	[0.283,0.285]	[0.402,0.405]
	[0.053,0.423]	[0.075,0.662]	[0.106,0.939]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Toothbrush category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML = 3$ to model observations of $T = 5$. Tuning parameter is set to be $\tau_n = 0.8$ and the number of bootstrap samples 30.

Table A17: DPO Estimates from the Toothbrush cateogry

brand	HBL	$HBL\|Y_t=j$	$HBL\|Y_t=Y_{t-1}=j$
YOPLAIT	[0.048,0.373]	[0.056,0.370]	[0.072,0.475]
	[0.000,0.411]	[0.000,0.417]	[0.000,0.534]
DANNON	[0.045,0.266]	[0.075,0.338]	[0.111,0.500]
	[0.000,0.304]	[0.000,0.424]	[0.000,0.626]
COLOMBO	[0.040,0.330]	[0.101,0.351]	[0.160,0.555]
	[0.000,0.378]	[0.000,0.432]	[0.000,0.683]

(a) Binary DPO Bounds

brand	HBL	$HBL\|Y_t=j$	$HBL\|Y_t=Y_{t-1}=j$
YOPLAIT	[0.038,0.526]	[0.044,0.691]	[0.000,0.823]
	[0.000,0.603]	[0.015,0.838]	[0.000,1.000]
DANNON	[0.038,0.447]	[0.053,0.895]	[0.000,1.000]
	[0.000,0.531]	[0.000,1.000]	[0.000,1.000]
COLOMBO	[0.031,0.426]	[0.063,0.840]	[0.000,0.966]
	[0.000,0.489]	[0.000,1.000]	[0.000,1.000]

(b) Multinomial DPO Bounds

This table summarizes the brand-specific HBL bounds estimated from the Yogurt category in the IRI Academic Datasets. Brands are ordered by popularity. For each brand, the first row reports estimated bounds and the second row reports 95% confidence region. Each household's last six purchases are reconstructed to be (a) a focal brand and a composite brand for the binary DPO model, and (b) the top brand, focal brand and a composite brand for the multinomial DPO model. Assumptions ST(1), MTS and MTR are imposed. The multinomial DPO model combines shorter models of length $ML=3$ to model observations of $T=5$. Tuning parameter is set to be $\tau_n=0.8$ and the number of bootstrap samples 30.

Table A18: DPO Estimates from the Yogurt cateogry

category	Likelihood Ratio (homogeneous)	Likelihood Ratio (GFE)
coffee	46.689	29.018
deodorant	118.639	110.629
mayonnaise	833.802	818.123
margarine	133.954	69.099
peanut butter	485.134	448.643
spaghetti sauce	109.754	108.661
toothbrushes	17.271	14.697
yogurt	5.046	29.679
critical value	3.841	7.815

This table summarizes the likelihood ratio tests using the estimation samples drawn from the IRI Academic Dataset. The null model is the static RUM with $\beta=0$ and the alternative model is the DDC with β allowed to vary freely. Critical values of the χ^2 statistic are for a the 5% significance level.

Table A19: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.563	0.181
price	-1.050	0.080	-1.048	0.080
loyalty	1.479	0.157	1.479	0.157
$FOLGERS_S$	0.549	0.383	0.519	0.383
$FOLGERS_L$	-0.913	0.315	-0.893	0.318
$MAXWELL_S$	0.840	0.348	0.819	0.344
$MAXWELL_L$	-0.988	0.344	-0.938	0.346
$PRIVATELABEL_S$	-0.900	0.405	-0.916	0.398

This table summarizes the static RUM and DDC coefficient estimates from the Coffee category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A20: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.990	0.103
price	-0.490	0.083	-0.498	0.081
loyalty	1.954	0.089	1.954	0.089
DEGREE	-4.530	0.299	-4.424	0.289
DOVE	-4.528	0.238	-4.413	0.234
MENNEN	-4.069	0.267	-4.111	0.258
OLDSPICE	-4.127	0.296	-4.091	0.280
RIGHT	-4.513	0.255	-4.444	0.261
SECRET	-4.152	0.367	-4.031	0.355

This table summarizes the static RUM and DDC coefficient estimates from the Deodorants category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A21: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.374	0.194
price	-0.714	0.144	-0.715	0.145
loyalty	3.207	0.152	3.207	0.150
ICANTBELIEVEITSNOTBUTTER	-3.780	0.351	-3.784	0.352
SMARTBALANCE	-3.887	0.357	-3.883	0.360
SHEDDSCOUNTRYCROCK	-4.493	0.300	-4.439	0.292
PROMISE	-4.025	0.469	-4.042	0.455

This table summarizes the static RUM and DDC coefficient estimates from the Margarine category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A22: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.996	0.042
price	-2.083	0.154	-2.077	0.155
loyalty	2.028	0.141	2.027	0.142
HELLMANNS	-1.347	0.305	-1.365	0.307
PRIVATELABEL	-2.860	0.245	-2.800	0.231
KRAFTMAYO	-2.283	0.254	-2.020	0.261

This table summarizes the static RUM and DDC coefficient estimates from the Mayonnaise category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A23: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	1.000	0.046
price	-0.861	0.084	-0.900	0.087
loyalty	1.617	0.047	1.617	0.047
JIF	-2.919	0.163	-2.906	0.165
PETERPAN	-3.809	0.147	-3.546	0.154
PRIVATELABEL	-3.419	0.148	-3.289	0.150
SKIPPY	-2.952	0.159	-2.900	0.164

This table summarizes the static RUM and DDC coefficient estimates from the Peanut Butter category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A24: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.929	0.206
price	-3.104	0.193	-3.103	0.192
loyalty	1.663	0.100	1.663	0.100
FRANCESCORINALDI	-1.914	0.203	-1.830	0.206
HUNTS	-2.724	0.236	-2.615	0.231
PREGO	-0.635	0.262	-0.621	0.264
RAGU	-0.085	0.271	-0.128	0.265

This table summarizes the static RUM and DDC coefficient estimates from the Spaghetti Sauce category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A25: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.478	0.208
price	-0.796	0.102	-0.792	0.101
loyalty	1.070	0.195	1.070	0.193
COLGATE	-2.995	0.395	-3.011	0.386
ORALB	-2.345	0.344	-2.333	0.348

This table summarizes the static RUM and DDC coefficient estimates from the Toothbrushes category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A26: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications

coefficient	MNL	s.e. (MNL)	DDC	s.e. (DDC)
beta	0.000	0.000	0.924	0.225
price	-0.830	0.083	-0.831	0.083
loyalty	1.993	0.120	1.993	0.120
YOPLAIT	-2.219	0.220	-2.251	0.219
DANNON	-2.691	0.184	-2.662	0.176
COLOMBO	-3.253	0.242	-3.147	0.249

This table summarizes the static RUM and DDC coefficient estimates from the Yogurt category in the IRI Academic Dataset. All specifications use the "group fixed-effects" approach to control for unobserved heterogeneity. Reported parameter values represent the population mean value of a parameter. Standard errors are computed via bootstrap.

Table A27: Likelihood Ratio Tests for the static and dynamic (DDC) RUM Specifications