

WORKING PAPER · NO. 2020-13

Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares

Jean-Pierre H. Dubé, Ali Hortaçsu, and Joonhwi Joo

FEBRUARY 2020

Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares
Jean-Pierre H. Dubé, Ali Hortaçsu, and Joonhwi Joo
February 2020
JEL No. D12,L00,L66,L81,M3

ABSTRACT

Although typically overlooked, many purchase datasets exhibit a high incidence of products with zero sales. We propose a new estimator for the Random-Coefficients Logit demand system for purchase datasets with zero-valued market shares. The identification of the demand parameters is based on a pairwise-differencing approach that constructs moment conditions based on differences in demand between pairs of products. The corresponding estimator corrects non-parametrically for the potential selection of the incidence of zeros on unobserved aspects of demand. The estimator also corrects for the potential endogeneity of marketing variables both in demand and in the selection propensities. Monte Carlo simulations show that our proposed estimator provides reliable small-sample inference both with and without selection-on-unobservables. In an empirical case study, the proposed estimator not only generates different demand estimates than approaches that ignore selection in the incidence of zero shares, it also generates better out-of-sample fit of observed retail contribution margins.

Jean-Pierre H. Dubé
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
jdube@chicagobooth.edu

Joonhwi Joo
800 W. Campbell Rd.
JSOM 13-326
Richardson, TX 75080
USA
joonhwi.joo@utdallas.edu

Ali Hortaçsu
Kenneth C. Griffin Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
hortacsu@uchicago.edu

1 Introduction

The seminal work by Berry (1994) and Berry et al. (1995) has generated a surge in empirical case studies that estimate discrete choice demand in the presence of endogenous marketing variables.¹ To obtain a viable aggregate demand prediction, the underlying model assumes that a market comprises a mass of consumers making discrete choices with non-zero choice probabilities for each of the available products. The combination of a continuum of consumers and non-zero choice probabilities is central to the estimation methodology. In turn, it also implies that each available product has a strictly positive market share.

In practice, many markets have products that fail to generate sales and exhibit zero market shares, even in the presence of a very large underlying consumer population. Conceptually, the observation of any products with a zero-valued market share rejects Berry et al. (1995)’s underlying model of demand. In practice researchers impose ad hoc modifications of the data themselves to force-fit them to exhibit strictly positive shares through product aggregation, time aggregation and/or selective trimming of the data sample to eliminate problematic observations. These ad hoc procedures are quite prevalent, often unbeknownst to the researcher. For instance, retail scanner data from IRI and Nielsen typically only report sales and marketing data for products that sold at least one unit in a given market. Force-fitting the model to these manipulated data potentially biases demand estimation, with adverse effects on estimated preferences and policy simulations based on the demand estimates. Of particular concern is the potential selection bias associated with a systematic modification or truncation of observations with zero shares.

To resolve the zero-market-share problem, we propose a new model of discrete choice demand that addresses the potential selection process determining whether a product generates a zero versus a positive market share. Our model includes a regime-switching process through which products are included or excluded from the choice set. This switching propensity serves as a selection correction procedure that allows each available product’s predicted demand to switch between a zero share and the strictly positive predicted share from the usual discrete choice model. While the paper focuses primarily on discrete-choice demand, we show that it is straightforward to apply the identification and estimation below to the CES demand model that is popular in the trade literature (e.g., Adao et al., 2017; Helpman et al., 2008).

The identification of our model primitives – the distribution of consumer preferences – relies on a novel modification of the moment conditions proposed in Berry et al. (1995). In addition to the usual concerns about the potential endogeneity of prices in consumers’ choices, we also account for the selection of market shares on unobserved shocks and we allow endogenous variables such as prices to enter into the selection equation. We rely on two sets of instruments. The first set comprises the usual instruments to correct for price endogeneity: instruments for prices that are excluded from both the consideration and purchase decisions. The second set contains new “consideration instruments” that

¹See for instance the survey by Nevo (2011), in economics, and Chintagunta et al. (2006), in marketing.

affect which products are effectively considered by consumers when they make their choices; but are excluded from the consumption utilities. These instruments could reflect consumer attention and the in-store search process, on the demand side, or availability and stock-outs, on the supply side. Following the literature on pairwise-difference estimation (e.g., Ahn and Powell, 1993; Blundell and Powell, 2004; Powell, 2001; Honoré and Powell, 2005; Aradillas-Lopez et al., 2007; Aradillas-Lopez, 2012), we then construct pairwise-difference moments that resolve the selection of product market shares on unobserved aspects of demand. To the best of our knowledge, this is the first application of pairwise-difference estimation in the context of demand estimation.

We then propose a three-step estimation procedure that allows for a fully non-parametric selection stage. In the first step, we estimate the reduced-form of prices, which we use to construct a control function (Blundell and Powell, 2004). Unlike Petrin and Train (2010), we only use the control function to address price endogeneity in the selection-stage, not in the consumers’ discrete choice stage. In the second stage, we nonparametrically estimate the product selection process using the control function to account for potential price endogeneity. Our nonparametric selection stage allows us to be agnostic about the exact mechanism determining which of the available products enter consumers’ choice sets and generate positive shares. In our third stage, we estimate the demand primitives of interest using a pairwise-differenced, weighted GMM estimator (PDWGMM) with nonparametric selection propensities as weights to resolve the selection problem.

The \sqrt{n} -consistency of our PDWGMM estimator follows from Aradillas-Lopez et al. (2007). We propose a bootstrap algorithm to conduct inference on the PDWGMM estimates. A series of Monte Carlo experiments verify the finite-sample performance of our estimator and the biases that arise from usual ad hoc data fixes to implement Berry et al. (1995)’s GMM estimator when it is subject to selection concerns. Conventional estimators that ignore selection exhibit biases that would lead to incorrect inferences about demand and preferences. They also confirm that PDWGMM produces consistent estimates comparable to conventional approaches when the data generating process does not exhibit selection on unobservables or when the data generating process does not produce zero market shares at all.

We apply our PDWGMM estimator to an empirical case study of the ice cream category using 18 months of daily, store-level point-of-sale data from a midwestern supermarket chain. A distinctive feature of these data is that we observe daily prices, unit costs, and feature advertising information even when a product does not sell.² A typical store carries around 400-500 unique ice cream UPCs on a given day; but over 60% of these products fail to sell at least one unit that day. Collapsing the time dimension to the monthly or quarterly level does not eliminate the zeros and could introduce an unintended aggregation bias due to time-varying marketing conditions (e.g., Gupta et al., 1996) In sum, weekly or even monthly frequency data do not resolve the zeros problem. Combining products into “items” by aggregating across pack sizes and flavors for a given brand would also eliminate some

²Most point-of-sale databases from providers like IRI and Nielsen have missing observations for products that do not sell in a given market and are typically aggregated to the weekly level.

of the zero-share issues. However, such aggregation would be impractical for the analysis of optimal product variety and consumer value creation.

We estimate the Random-Coefficients Logit demand for ice cream using our PDWGMM estimator. We compare these results to those obtained from the conventional GMM estimator of the Random-Coefficients Logit demand model fit to data that either drop observations with zero market share or impute a numerically small value. We find that PDWGMM generates considerably more price elastic estimates of demand than alternative, ad hoc solutions to the zero-market-share problem.

Finally, we exploit product margin data to construct a specification test of our PDWGMM estimator that accounts for the potential selection bias associated with products generating a zero-valued market share. Our test consists of predicting the retailer’s product-specific margins using our demand and elasticity estimates under various assumptions about pricing conduct. For each of the conduct assumptions, we find that PDWGMM generates a much better fit of the observed margins than ad hoc approaches.

The model and estimation procedure contribute to the vast extant literature on aggregate discrete-choice demand estimation by accommodating zero shares. In related work, Gandhi et al. (2019) propose a different modeling solution that treats the average purchase probabilities as a missing data problem and market shares as an imperfect proxy. In their model, zero market shares arise from sampling error due to a small consumer population. They formulate a moment inequalities estimator that generates bounds for the preference parameters of interest. However, zero shares are often observed even in very large markets with tens of thousands of underlying consumers, such as large chain-store supermarkets and mass merchandisers. We propose an approach that allows for zero-valued market shares even with a very large underlying consumer population. We also use the product margin data to construct a model specification test.

An advantage of our nonparametric selection stage is that we remain agnostic about the mechanism causing some products to generate zero share of sales. This mechanism could be limited consideration by consumers, or stock-outs/unavailability of the product on the shelf. Our consideration instruments connect our approach to the literature on consumer consideration sets whereby consumers often consider only a small portion of the available products when making purchase decisions (e.g., Hauser, 1978; Ratchford, 1980; Shugan, 1980; Hauser and Wernerfelt, 1990; Moorthy et al., 1997; Kim et al., 2010; Honka, 2014; Bronnenberg et al., forthcoming). Integrating the additional structure through which consumers search for products and form consideration sets typically requires individual-level data on consumers’ heterogeneous consideration sets (e.g., Honka, 2014) or at least additional moments on the search ranks in addition to market shares conditional on search (Kim et al. (2017)). In addition to data, these approaches require additional conduct assumptions about consumer search and attention. A recent exception is Abaluck and Adams (2018) which exploits the Slutsky symmetry property of full-information discrete choice demand models to identify a test for limited consideration. However, their approach cannot account for the zero market shares

encountered in aggregate data sets and their estimator would not accommodate endogenous regressors. Since our selection process is agnostic, our work also relates to demand estimation with data that aggregates market shares across stores and time periods with different “choice sets” due, for instance, to differing promotional conditions or product availabilities (e.g., Tenn, 2006; Bruno and Vilcassim, 2008) and stock-outs (e.g., Conlon and Mortimer, 2013).

Our work is also related to Helpman et al. (2008)’s CES demand estimation framework for bilateral trades when zero trade shares are observed, albeit with a fully parametric selection stage that does not allow for endogenous covariates during the selection stage and without random coefficients. We add to this work by developing a nonparametric selection stage that potentially exhibits price endogeneity, and by allowing for random coefficients in the demand model.

The remainder of this paper is organized as follows. Section 2 derives the model of demand, consideration and pricing. Sections 3 and 4 present the identification and estimation of the model, respectively. Section 5 presents the results from a Monte Carlo study. Empirical results are presented in section 6. Section 7 shows the extension of our framework to the CES demand system.

2 Model

2.1 The Random-Coefficients Logit Demand System

We briefly discuss the derivation of the Random-Coefficients Logit demand system used extensively in the aggregate demand estimation literature. Consumers make purchases from the set of available substitute products $\mathcal{J} = \{1, \dots, J + 1\}$ with corresponding prices $\mathbf{p} = (p_1, \dots, p_{J+1})'$ and where the $J + 1$ product is a “no purchase” or “outside” option with price $p_{J+1} = 0$. Consumer h spends her budget y to maximize a utility function $U(\mathbf{q}, y; \boldsymbol{\theta}^h)$ defined over her consumption choices, $\mathbf{q} = (q_1, \dots, q_{J+1})'$, and expenditure on an essential numeraire good, q_0 :

$$U(\mathbf{q}, y; \boldsymbol{\theta}^h) = \sum_{j=1}^{J+1} (\chi_j^h q_j) \exp(\alpha^h q_0)$$

where χ_j^h is household h 's perceived quality of product j and α^h is the preference for the essential numeraire good. Assume perceived quality consists of a deterministic and a stochastic component: $\chi_j^h = \exp(\mathbf{x}_j' \boldsymbol{\beta}^h + \xi_j + \epsilon_j^h)$ and $\chi_{J+1}^h = \exp(\epsilon_{J+1}^h)$, where $\epsilon_j^h \sim \text{i.i.d. EV}(0, 1) \forall j$, \mathbf{x}_j is a vector of exogenous product attributes, ξ_j is a scalar vertical characteristic that is unobserved by the researcher and $\boldsymbol{\beta}^h$ are taste parameters for the characteristics. When consumption is discrete, $q_j \in \{0, 1\} \forall j$, consumer h faces the following choice-specific indirect utilities (in logs):

$$\begin{aligned} u_j^h &= \alpha^h (y - p_j) + \mathbf{x}_j' \boldsymbol{\beta}^h + \xi_j + \epsilon_j^h \\ &\vdots \\ u_{J+1}^h &= \alpha^h y + \epsilon_{J+1}^h. \end{aligned} \tag{2.1}$$

We assume household h 's preferences $\boldsymbol{\theta}^h := \begin{pmatrix} \alpha^h \\ \boldsymbol{\beta}^h \end{pmatrix}$ follows a distribution $F_{\boldsymbol{\theta}}(\boldsymbol{\theta}^h; \boldsymbol{\Theta})$ parameterized by $\boldsymbol{\Theta}$.

The household-specific mean utility of alternative j becomes:

$$\bar{u}_j^h \equiv \bar{u}_j^h(\alpha^h, \boldsymbol{\beta}^h) = \alpha^h (y - p_j) + \mathbf{x}'_j \boldsymbol{\beta}^h + \xi_j. \quad (2.2)$$

The predicted market share for product j is then

$$\begin{aligned} \pi_j(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \boldsymbol{\Theta}) &= \int \int_{\{\boldsymbol{\theta}^h, \epsilon_j^h | u_j^h \geq u_k^h, \forall k \neq j\}} dF_{\epsilon}(\epsilon^h) dF_{\boldsymbol{\theta}}(\boldsymbol{\theta}^h; \boldsymbol{\Theta}) \\ &= \int \frac{\exp(\alpha^h p_j + \mathbf{x}'_j \boldsymbol{\beta}^h + \xi_j)}{1 + \sum_{k=1}^J \exp(\alpha^h p_k + \mathbf{x}'_k \boldsymbol{\beta}^h + \xi_k)} dF_{\boldsymbol{\theta}}(\boldsymbol{\theta}^h; \boldsymbol{\Theta}). \end{aligned} \quad (2.3)$$

An important feature of this model is that the predicted market shares must be strictly positive: $\pi_j(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \boldsymbol{\Theta}) > 0, \forall j$.

2.2 Product Selection

We assume that the “no purchase” option $J + 1$ always enters consumers' choice sets. However, for each of the $j = 1, \dots, J$ products, we let $d_j \in \{0, 1\}$ indicate whether product j enters the choice sets or not. Let $\mathcal{C} = \{j \in \mathcal{J} | d_j = 1 \text{ or } j = J + 1\}$ denote those products that enter consumers' choice sets. To model the selection process through which products enter consumers' choice sets, let

$$d_j = \mathbf{1}(\varphi(\mathbf{w}_j) + \eta_j > 0), \quad j = 1, \dots, J \quad (2.4)$$

where \mathbf{w}_j is a vector of observed (to the researcher) selection-generating factors and $\eta_j \sim G_{\eta}(\cdot)$ is a random selection shock with continuous and strictly increasing distribution $G_{\eta}(\cdot)$. In principle, \mathbf{w}_j could include product attributes and prices for product j as well as its competitors.

We focus on the case where the regime-switching process selects a product out of the choice set, $d_j = 0$:

$$\pi_j(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \boldsymbol{\Theta}) = \begin{cases} \int \frac{\exp(\alpha^h p_j + \mathbf{x}'_j \boldsymbol{\beta}^h + \xi_j)}{1 + \sum_{k \in \mathcal{C}} \exp(\alpha^h p_k + \mathbf{x}'_k \boldsymbol{\beta}^h + \xi_k)} dF_{\boldsymbol{\theta}}(\boldsymbol{\theta}^h; \boldsymbol{\Theta}), & d_j = 1 \\ 0, & d_j = 0 \end{cases}. \quad (2.5)$$

As we discuss in section 3 below, product selection complicates the estimation of the demand system (2.3), especially when demand shocks are correlated with selection shocks: $\text{cov}(\xi_j, \eta_j) \neq 0$.

2.3 Pricing

A long extant literature on demand estimation has analyzed the role of firms that strategically condition their prices on aspects of demand that are unobserved to the researcher (e.g., Berry,

1994; Berry et al., 1995; Besanko et al., 1998). Suppose that the reduced-form of equilibrium prices in the market can be written as:

$$p_j = \Pi(\mathbf{r}_j; \boldsymbol{\varrho}) + \nu_j \quad (2.6)$$

where $\mathbf{r}_j = (\tilde{\mathbf{x}}_j, \tilde{\mathbf{z}}_j)$ and $\tilde{\mathbf{z}}_j$ are supply-shifting factors that influence firms' marginal costs and are excluded from the choice-specific utilities (2.1), and $\nu_j \sim F_\nu(\cdot)$ is a random supply shock. As we discuss in section 3.1 below, strategic pricing can also complicate the estimation of the demand system (2.3) when $\text{cov}(\nu_j, \xi_j) \neq 0$, or $\text{cov}(\nu_j, \eta_j) \neq 0$, or both.

3 Identification

3.1 BLP's Moment Conditions and Zero Shares

We now discuss the estimation of the demand system (2.3) from section (2.1). At least since Berry (1994) and Berry et al. (1995), the extant literature has focused on demand estimation in the presence of endogenous prices: $\text{cov}(p_j, \xi_j) \neq 0$.

From the model section, we can construct moment conditions:

$$\mathbb{E}[\xi_j | \mathbf{z}_j] = 0 \quad (3.1)$$

where as before $\mathbf{z}_j = (\mathbf{x}_j, \tilde{\mathbf{z}}_j)$ are the price instruments. We can then construct the empirical analog of these moment conditions. Let $\mathbf{s} = (s_1, \dots, s_{J+1})$ be the vector of observed market shares. Using the demand model (2.3), we can define the following mapping between the demand shocks $\boldsymbol{\xi}$ and the observed market shares:

$$s_j = \pi_j(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \boldsymbol{\Theta}). \quad (3.2)$$

Berry (1994) proves that the market share equations in (3.2) are invertible in the demand shocks: $\boldsymbol{\xi}(\boldsymbol{\Theta}) = \boldsymbol{\pi}^{-1}(\mathbf{x}, \mathbf{s}; \boldsymbol{\Theta})$. Berry et al. (1995) propose a GMM estimator using the empirical moments:

$$\mathbb{E}[\xi_j(\boldsymbol{\Theta}) | \mathbf{z}_j] = 0. \quad (3.3)$$

If $s_j = 0$ even for only one product, we reject the demand model in section (2.1) and the inversion procedure used to construct empirical moments is no longer valid. In practice, many micro datasets contain observations with zero market shares even when a product was available for sale, leading to the alternative formulation of demand in (2.5). The incidence of zero shares is not a new problem. As discussed in Cooper and Nakanishi (1988), most researchers manually alter the observed data ad hoc to force all the shares to be strictly positive. Two ad hoc data modifications are typically used in practice: (i) impute an arbitrarily small value whenever $d_j = 0$ (e.g., $s_j = 1.e^{-12}$ if $d_j = 0$); or

(ii) drop all observations with $d_j = 0$ (i.e., assume $j \notin \mathcal{J}$).³ We expect approach (i) to suffer from a bias that is sensitive to the specific “small number” used for the imputation.

Recently, Gandhi et al. (2019) proposed a modified version of the imputation procedure. They start with the assumption that zero-valued market shares arise from small markets with too few consumers for the market shares, s_j , to provide a reliable approximation of the expected unconditional choice probabilities, π_j , from the discrete choice model. They derive lower and upper bounds for π_j using s_j . These bounds can accommodate observations where $s_j = 0$ by introducing a Laplace “correction factor” that forces $s_j \in (0, 1)$. Standard bounds estimation techniques are then applied to construct an estimator (e.g., Manski and Tamer, 2002; Imbens and Manski, 2004; Chernozhukov et al., 2007; Ciliberto and Tamer, 2009).

We are more concerned with approach (ii), which is much more commonly used in practice.⁴ Using the selection model in section 2.2, approach (ii) consists of using the following moment condition:

$$\mathbb{E}[\xi_j | \mathbf{z}_j, d_j = 1] = 0. \quad (3.4)$$

Moment condition (3.4) may hold if there is no selection on unobservable product characteristics, $\text{cov}(\xi_j, \eta_j) = 0$, implying that the dropped observations with $d_j = 0$ are “missing at random.” However, if $\text{cov}(\xi_j, \eta_j) \neq 0$, we are concerned with the potential for selection bias associated with dropping observations with zero shares: $\mathbb{E}[\xi_j | \mathbf{z}_j, d_j = 1] \neq 0$. In this case, GMM estimation using (3.4) will not produce consistent estimates of Θ .

3.2 Accommodating Zero-Valued Shares

We now derive a new set of moment conditions that can be used to obtain consistent GMM estimates of Θ with zero market shares and $\text{cov}(\xi_j, \eta_j) \neq 0$. The estimator also accommodates the endogeneity bias that arises when prices are correlated with both demand shocks and selection shocks: $\text{cov}(p_j, \xi_j) \neq 0$ and $\text{cov}(p_j, \eta_j) \neq 0$. Our solution consists of using moment conditions based on differences, as in the pairwise-difference estimation literature (e.g., Ahn and Powell, 1993; Blundell and Powell, 2004; Powell, 2001; Honoré and Powell, 2005; Aradillas-Lopez et al., 2007; Aradillas-Lopez, 2012).

³Researchers frequently use yet another approach whereby each of the products is aggregated up to the brand level, often combining dozens of different products (e.g., different pack sizes, flavors etc) into a composite “brand” alternative (e.g., Nevo, 2001).

⁴Most CPG datasets only report market shares for products in a given market that had a positive quantity sold. For instance, data suppliers like Nielsen and IRI Marketing receive quantity and revenue data from retail partners and, therefore, cannot determine the average price for a product that did not have any sales. Consequently, researchers using such scanner data implicitly use assumption (ii).

3.2.1 Pairwise-Difference Moment Condition

Let μ_j denote the propensity score for alternative j not being selected into the choice set. We assume that μ_j is continuously distributed and that there exists an unknown, smooth, strictly increasing function $\iota(\cdot)$ such that we can re-write the demand shocks as follows:

$$\xi_j = \iota(\mu_j) + \zeta_j, \quad (3.5)$$

with $\mathbb{E}[\zeta_j | \mu_j, \mathbf{z}_j] = 0$. Even if $\mathbb{E}[\xi_j | \mathbf{z}_j, d_j = 1] \neq 0$, we can construct an alternative moment condition based on differences between the demand shocks of two goods i and j :

$$\mathbb{E}[\xi_i - \xi_j | \mathbf{z}_i, \mathbf{z}_j, \mu_i = \mu_j, d_i = d_j = 1] = 0. \quad (3.6)$$

The use of a moment based on differences follows from the pairwise-difference estimation literature (e.g., Ahn and Powell, 1993; Blundell and Powell, 2004; Powell, 2001; Honoré and Powell, 2005; Aradillas-Lopez et al., 2007; Aradillas-Lopez, 2012).

To illustrate, consider the Homogeneous Logit model of demand which has a degenerate distribution of random coefficients. Using (3.5), we can invert the demand model as follows:

$$\ln\left(\frac{s_j}{s_0}\right) = \alpha p_j + \mathbf{x}'_j \boldsymbol{\beta} + \iota(\mu_j) + \zeta_j.$$

Now consider a pair of products i and j with the same non-selection propensities, $\mu_i = \mu_j$, and hence $\iota(\mu_i) = \iota(\mu_j)$:

$$\begin{aligned} \ln\left(\frac{s_i}{s_0}\right) &= \alpha p_i + \mathbf{x}'_i \boldsymbol{\beta} + \iota(\mu_i) + \zeta_i \\ \ln\left(\frac{s_j}{s_0}\right) &= \alpha p_j + \mathbf{x}'_j \boldsymbol{\beta} + \iota(\mu_j) + \zeta_j. \end{aligned}$$

Taking differences yields:

$$\ln\left(\frac{s_i}{s_0}\right) - \ln\left(\frac{s_j}{s_0}\right) = \alpha(p_i - p_j) + (\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\beta} + (\zeta_i - \zeta_j).$$

We can then construct the following conditional moment for estimation purposes:

$$\begin{aligned} 0 &= \mathbb{E}[\zeta_i - \zeta_j | \mathbf{z}_i, \mathbf{z}_j, \mu_i = \mu_j, d_i = d_j = 1] \\ &= \mathbb{E}[\xi_i - \xi_j | \mathbf{z}_i, \mathbf{z}_j, \mu_i = \mu_j, d_i = d_j = 1]. \end{aligned} \quad (3.7)$$

An obvious limitation of the pairwise-difference formulation is that the intercept term of the indirect utility function is not identified. Since the intercept term is typically a nuisance term in most applications, this will not prevent the researcher from predicting market shares and conducting

policy experiments with the estimated model. In sum, the levels of $\hat{\xi}_j$ contain both the absorbed intercept term and the factors driving selection on unobservables. It is not possible to disentangle these two components separately.

In the next section, we discuss instruments that vary the propensities, μ_i and μ_j , independently of preferences.

3.2.2 Non-Selection Propensities

The identification of the propensities μ_j relies on consideration instruments that are excluded from the choice-specific utilities (2.1) but are allowed to be correlated with the demand shocks ξ_j through $\text{cov}(\xi_j, \eta_j)$. We define the propensity score μ_j as the expectation of the selection variable $1 - d_j$ conditional on the observed prices p_j , product attributes and price instruments \mathbf{z}_j , and consideration instruments $\tilde{\mathbf{w}}_j$:

$$\mu_j = 1 - \mathbb{E}[d_j | p_j, \mathbf{z}_j, \tilde{\mathbf{w}}_j]. \quad (3.8)$$

The consideration instruments, $\tilde{\mathbf{w}}_j$, could include promotions such as feature advertising and in-store displays which are known to affect demand, but are unlikely to generate *consumption value* (e.g., Mehta et al., 2003).

A concern that has not previously been considered in the extant demand estimation literature (i.e., $\mathcal{C} = \mathcal{J}$) is the potential for price endogeneity during the selection process if $\text{cov}(\nu_j, \eta_j) \neq 0$ causing p_j to correlate with η_j . Our solution consists of constructing a control variable ν_j from the pricing equation such that the moment condition (3.6) is valid with $\mathbf{w}_j = (\nu_j, \tilde{\mathbf{w}}_j)$, $\mu_j = \mu(\mathbf{w}_j)$ where $\mu(\mathbf{w}_j)$ is defined by:

$$\mu(\mathbf{w}_j) := 1 - \mathbb{E}[d_j | \mathbf{w}_j] = 1 - \mathbb{E}[\mathbf{1}(\varphi(\mathbf{w}_j) + \eta_j > 0) | \mathbf{w}_j] = 1 - G_\eta(-\varphi(\mathbf{w}_j)) \quad (3.9)$$

where $\mu(\mathbf{w}_j)$ is the (reduced-form) conditional probability of product j not being included in the subset of alternatives being selected and $\varphi(\cdot)$ is an unknown function. We can then construct a control function using the assumption $\mathbb{E}[\nu_j | \mathbf{z}_j] = 0$ where $\tilde{\mathbf{z}}_j$ is excluded from the non-selection propensity.

Finally, for nonparametric identification of the non-selection propensity μ_j , we assume the following distributional exclusion restriction on the conditional distribution of η_j (e.g., Blundell and Powell, 2004)

$$\Pr(\eta_j < \eta | p_j, \mathbf{z}_j, \tilde{\mathbf{w}}_j) = \Pr(\eta_j < \eta | \nu_j, \tilde{\mathbf{w}}_j) \equiv G_\eta(-\varphi(\mathbf{w}_j)) \quad \forall \eta \in \mathbb{R}. \quad (3.10)$$

The assumption requires any dependence of η_j on prices p_j , exogenous product attributes and/or price instruments \mathbf{z}_j to be characterized by the control function ν_j that comes from the pricing equation (2.6) in section 2.3. Using this distributional exclusion restriction assumption with moderate regularity conditions, we can identify $G_\eta(\cdot)$ and $\varphi(\cdot)$, and thus, the non-selection propensity $\mu(\mathbf{w}_j)$

nonparametrically (see Matzkin (1992) for a proof). A conceptual advantage of our nonparametric approach is that we allow $G_\eta(\cdot)$ and $\varphi(\cdot)$ to be very flexible as they do not need to depend on a finite-dimensional parameter vector, such as γ in $\varphi(\mathbf{w}_j) = \mathbf{w}'_j\gamma$, if one adopts a parametric or semiparametric specification of the selection equation. Thus, our approach is agnostic about the exact mechanism through which the subset of alternatives being selected is formed. In the next section, we propose a three-step estimator of the Random-Coefficients Logit model that is robust to zero shares (selection) and endogenous prices.

4 A Three-Step Estimator with a Nonparametric Selection Stage

We now propose a three-step estimator using the pairwise-difference moment conditions in equation (3.6). The first step estimates the reduced form of prices. We use the reduced form to construct a control function that is then used in a second step to obtain a nonparametric estimate of our binary selection process. Finally, the third step estimates the structural parameters of interest, Θ , using a constrained optimization procedure that fits our “differences” moments subject to the selection process. The approach is related to pairwise-differencing estimation of selection models with a nonparametric first-stage (see for instance Ahn and Powell, 1993, who use a linear second stage). Aradillas-Lopez et al. (2007); Aradillas-Lopez (2012) extend Ahn and Powell (1993); Honoré and Powell (2005)’s setup to allow for a non-parametric first-stage propensity score and a non-linear second stage.

4.1 Stage 1: Pricing Equation

Assume that the pricing equation has the following structure:

$$p_j = \Pi(\mathbf{r}_j; \boldsymbol{\varrho}) + \nu_j, \quad (4.1)$$

where the functional form of $\Pi(\mathbf{r}_j; \boldsymbol{\varrho})$ is known to the researcher. Suppose $\hat{\boldsymbol{\varrho}}$ is a \sqrt{n} -consistent estimator of $\boldsymbol{\varrho}$. The residual, which will be used as the control variable in the second stage of estimation, is defined as follows:

$$\hat{\nu}_j = p_j - \Pi(\mathbf{r}_j; \hat{\boldsymbol{\varrho}}). \quad (4.2)$$

5

As in section 3.2.2, we construct an instrument matrix \mathbf{W} with j^{th} row $\mathbf{w}'_j := \begin{pmatrix} \hat{\nu}_j & \tilde{\mathbf{w}}'_j \end{pmatrix}$. As explained above in section 2.2, the selection process could be influenced by prices and product characteristics. However, under our distributional exclusion restriction assumption (3.10), we do not include p_j or \mathbf{x}_j in \mathbf{w}_j because $G_\eta(\eta|p_j, \mathbf{x}_j, \nu_j, \tilde{\mathbf{w}}_j) = G_\eta(\eta|\nu_j, \tilde{\mathbf{w}}_j)$, $\forall \eta \in \mathbb{R}$. Blundell and Powell

⁵In theory, one could estimate $\Pi(\mathbf{r}_j; \boldsymbol{\varrho})$ nonparametrically. However, a nonparametric estimator will typically converge at a slower rate than \sqrt{n} . The slower rate of convergence could, in turn, affect the asymptotic properties of the second and third stage estimators which might adversely affect our bootstrap procedure for inference.

(2004) show that, under this distributional exclusion restriction, the residual $\hat{\nu}_j \equiv p_j - \Pi(\mathbf{r}_j; \hat{\boldsymbol{\rho}})$ can be used as a control variable in estimating (2.4). Recall that the control function is used only to allow for potential price endogeneity in the estimation of the consideration selection probabilities.⁶ When we estimate the market share functions, we will use instrumental variables to correct for the endogeneity of prices.

4.2 Stage 2: Selection Stage – Nonparametric Estimation of the Control Function

Recall that our selection model for product j is given by:

$$\mu(\mathbf{w}_j) = 1 - \mathbb{E}[\mathbf{1}(\varphi(\mathbf{w}_j) + \eta_j > 0) | \mathbf{w}_j]. \quad (4.3)$$

In practice, researchers using aggregate market share data observe very little about the selection process other than the incidence of zero shares. Consequently, the functional form of $\varphi(\cdot)$ is unknown. If we assume that η_j has an everywhere-increasing continuous distribution, we can estimate μ_j nonparametrically for each observation j using the Nadaraya-Watson kernel regression estimator:

$$\hat{\mu}_j \equiv \hat{\mu}(\mathbf{w}_j) = 1 - \frac{\sum_k d_j H\left(\frac{\mathbf{w}_k - \mathbf{w}_j}{b_n}\right)}{\sum_k H\left(\frac{\mathbf{w}_k - \mathbf{w}_j}{b_n}\right)}, \quad (4.4)$$

where $H(\cdot)$ is a multivariate kernel function, and b_n is a bandwidth sequence for a sample of size n .⁷ In principle, one could impose more structure on (4.3) if more data was available about the specific selection mechanism, such as moments regarding consumers' consideration sets or moments regarding stock-outs and changes in availability.

4.3 Stage 3: Random-Coefficients Estimator with Nonparametric Selection Correction

The third stage of the estimator adapts the pairwise-differencing estimation procedure with nonparametric selection correction by Ahn and Powell (1993); Aradillas-Lopez et al. (2007); Aradillas-Lopez (2012) to random-coefficient demand estimation. We refer the interested reader to Ahn and Powell (1993); Aradillas-Lopez et al. (2007); Aradillas-Lopez (2012) for a detailed discussion of alternative estimation procedures (using bias-reducing higher-order kernel, trimming sequences or jackknife procedures) and for the proof of the \sqrt{n} -consistency and asymptotic normality of the estimator.

⁶Our application of a control function is different from Petrin and Train (2010). A crucial difference is that Petrin and Train (2010) include the control function from the pricing equation directly into the choice-specific utility. In contrast, the control function ν_j in our framework enters only the selection stage to correct for the endogeneity in the selection stage, not the choice-specific utility directly. In addition, unlike Petrin and Train (2010), we do not specify the parametric functional form of the unobservables.

⁷This bandwidth sequence b_n can differ for each element of \mathbf{w}_j if one chooses to use a product kernel.

We first define a differencing operator Δ . If Δ is applied to a pair of indexed vectors or scalars, it denotes the difference of two (column) vectors/scalars. For example, $\Delta \mathbf{z}_{i,j} := \mathbf{z}_i - \mathbf{z}_j$. If Δ is applied to a (column) vector or matrix without an index, it is a (column) vector/matrix of a collection of pairwise-differenced rows. For example, suppose \mathbf{Z} is an $n \times p$ matrix. Then, $\Delta \mathbf{Z}$ is the $\begin{pmatrix} n \\ 2 \end{pmatrix} \times p$ matrix where the first row of $\Delta \mathbf{Z}$ is $\Delta \mathbf{z}'_{1,2}$, second row is $\Delta \mathbf{z}'_{1,3}$, and so on.

Our proposed pairwise-differencing estimator uses the sample analogue of the following moment condition:

$$\mathbb{E} [\Delta \mathbf{z}_{i,j} \Delta \xi_{i,j} | \mu_i = \mu_j, d_i = d_j = 1] = \mathbf{0} \quad (4.5)$$

which is implied by (3.6).⁸ This moment condition resolves the selection problem, as illustrated for the case of Homogeneous Logit demand in equation (3.7) of section 3.2.2.

Unfortunately, as is common in the pairwise-differencing estimators, the empirical analog of the population moment condition (4.5) is complicated by the fact that μ_i exactly equals μ_j with probability 0 because μ_i is assumed to be continuously distributed. In addition, we only observe $\hat{\mu}_i$, the consistent estimate of μ_i from the previous stage. Therefore, we replace μ_i with its prediction $\hat{\mu}_i$, and use kernel weights that assign more weight to pairs of observations i and j when $\hat{\mu}_i$ and $\hat{\mu}_j$ are close. Define the kernel weights $\hat{\omega}_{i,j}$ by:

$$\hat{\omega}_{i,j} = \frac{1}{h_{n_d}} \kappa \left(\frac{\hat{\mu}_i - \hat{\mu}_j}{h_{n_d}} \right), \quad (4.6)$$

where $\kappa(\cdot)$ is a kernel function and h_{n_d} is a bandwidth sequence for a sample of size n_d , where $h_{n_d} \rightarrow 0$ as $n_d \rightarrow \infty$. The estimator corresponding to the moment condition (4.5) is a weighted GMM estimator with weights $\hat{\omega}_{i,j}$.

Our third-stage estimator is similar to the usual GMM estimator for the usual Random-Coefficients Logit demand model. Let Φ be a GMM weighting matrix constructed from the pairwise-differenced instrument matrix $\Delta \mathbf{Z}$. Also let $\mathbf{g}_{i,j}(\Theta) = \Delta \mathbf{z}_{i,j} \Delta \xi_{i,j}(\Theta)$ be the sample analog of our moment conditions (4.5). Then, with a slight modification of Ahn and Powell (1993), the pairwise-difference weighted GMM estimator is defined as the solution to the following constrained optimization prob-

⁸A sufficient condition for

$$\mathbb{E} [(\mathbf{z}_i - \mathbf{z}_j) (\xi_i - \xi_j) | \mu_i = \mu_j, d_i = d_j = 1] = 0$$

is that $\mathbb{E} [\mathbf{z}_i \xi_i - \mathbf{z}_j \xi_j | \mu_i = \mu_j, d_i = d_j = 1] = 0$ and $\mathbb{E} [\mathbf{z}_j \xi_j - \mathbf{z}_i \xi_i | \mu_i = \mu_j, d_i = d_j = 1] = 0$. Recalling (3.6), we can integrate over \mathbf{z}_i to obtain

$$\mathbb{E} [\xi_i - \xi_j | \mathbf{z}_i, \mathbf{z}_j, \mu_i = \mu_j, d_i = d_j = 1] = 0$$

which yields

$$\mathbb{E} [\xi_i - \xi_j | \mathbf{z}_j, \mu_i = \mu_j, d_i = d_j = 1] = 0,$$

and, in turn, implies

$$\mathbb{E} [(\xi_i - \xi_j) \mathbf{z}_j | \mu_i = \mu_j, d_i = d_j = 1] = \mathbf{0}.$$

Showing $\mathbb{E} [(\xi_i - \xi_j) \mathbf{z}_i | \mu_i = \mu_j, d_i = d_j = 1] = \mathbf{0}$ is similar.

lem:

$$\Theta^{PDWGMM} = \arg \min_{(\Theta, \xi)} \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{\omega}_{i,j} \mathbf{g}_{i,j}(\Theta)' \Phi \mathbf{g}_{i,j}(\Theta) \quad (4.7)$$

$$s.t. \quad \mathbf{s} = \boldsymbol{\pi}(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \Theta) \quad (4.8)$$

where $\hat{\omega}_{i,j}$ are the weights, pre-computed from the second stage according to (4.6). (4.8) is the market share equations for the observations with $s_j > 0$, as in equation (2.3). In addition to the standard instruments, we also include the Gandhi and Houde (2019)'s *product differentiation* instruments. Gandhi and Houde (2019) show that the inclusion of these additional instruments improves both the precision and numerical stability of the variance terms for the random coefficients. Appendix A provides specific details on the MPEC formulation (Dubé et al., 2012) of the estimation problem defined by (4.7)-(4.8). The appendix also describes the exact instrument matrix, \mathbf{Z} , and GMM weighting matrix, Φ , used during estimation.

The following proposition, which we adapt from Aradillas-Lopez et al. (2007), establishes the \sqrt{n} -consistency and asymptotic normality of the estimator proposed in (4.7)-(4.8). In addition to the usual regularity conditions for a nonparametric estimator such as on the kernel choice or bandwidth sequence stated in Aradillas-Lopez et al. (2007), the proposition imposes the following assumptions that were explained above: (i) continuity of at least one element of the exogenous product characteristics, (ii) the distributional exclusion restriction (3.10), and (iii) the \sqrt{n} -consistency of the pricing equation parameter estimates.

Proposition. *Consider the following assumptions:*

- (i) \mathbf{x}_j contains at least one continuous element.
- (ii) $G_\eta(\eta|p_j, \mathbf{x}_j, \nu_j, \tilde{\mathbf{w}}_j) = G_\eta(\eta|\nu_j, \tilde{\mathbf{w}}_j)$, $\forall \eta \in \mathbb{R}$.
- (iii) The functional form of the pricing equation $\Pi(\mathbf{r}_j; \boldsymbol{\rho})$ in (2.6) is known to the researcher, and the estimator $\hat{\boldsymbol{\rho}}$ for $\boldsymbol{\rho}$ is \sqrt{n} -consistent.

If assumptions (i)-(iii) hold along with several regularity conditions (Assumptions 1-17 of Aradillas-Lopez et al. (2007)), then, the estimator Θ^{PDWGMM} defined in (4.7)-(4.8) is \sqrt{n} -consistent and asymptotically normal.

Proof. See Aradillas-Lopez et al. (2007). □

Appendix A.2 provides specific details on our implementation of the MPEC estimation problem. Using the MPEC estimator in (4.7) for inference is complicated because it conditions on predictions from Stages 1 and 2. Appendix A.3 describes a bootstrap procedure for conducting inference using Θ^{PDWGMM} . Since bootstrapped Random-Coefficients Logit demand estimation is new in the literature, the appendix also derives a bootstrap approach for the case when all the observed market

shares are strictly positive. This scenario could be of interest if, for example, some of the observed product characteristics or instruments are predictions from a pre-estimation stage.

5 A Monte Carlo Study

5.1 Data Generating Process

We now study the finite-sample performance of our PDWGMM estimator both in sampling experiments with selection bias on zero shares and without. In each of our Monte Carlo experiments, we compare PDWGMM to the commonly-used Drop-Zero estimator, and Impute-Zero estimator. We consider three different data-generating processes: (a) the incidence of zero shares is selected on unobserved aspects of demand, (b) the incidence of zero shares is not selected on unobserved aspects of demand, and (c) strictly positive market shares. Scenarios (a) and (b) mimic typical, real-world consumer goods datasets where we observe many products that fail to sell any units. Of interest is whether our PDWGMM estimator provides reliable estimates regardless of whether the incidence of zero shares is selected on unobserved aspects of demand. Scenario (c) represents the data-generating process implied by the model in Berry et al. (1995) with a continuum of consumers generating strictly positive market shares. In this scenario, the PDWGMM estimator should be consistent and we anticipate it to provide comparable estimates to the Drop-Zero estimator (i.e., Berry et al. (1995)'s GMM estimator).

Each of our Monte Carlo experiments consists of $T = 100$ markets, each with $J = 100$ available products. We assume a linear choice-specific utility for each product:

$$u_{jt}^h = \beta_0^h + \alpha^h p_{jt} + \sum_{k=1}^3 x_{jkt} \beta_k^h + \xi_{jt} + \epsilon_{jt}^h$$

where $x_{jkt} \sim$ i.i.d. $U[0, 1]$ and $\epsilon_{jt}^h \sim$ i.i.d. $EV(0, 1)$. For the taste parameters, we assume $\alpha^h \sim \mathcal{N}(-2, 1)$, $\beta_0^h \sim \mathcal{N}(2, 0.25)$, $\beta_1^h \sim \mathcal{N}(1, 1)$, $\beta_2^h \sim \mathcal{N}(2, 4)$, and $\beta_3^h \sim \mathcal{N}(2, 4)$. As explained above, the intercept β_0^h is not separately identified for our pairwise-differencing estimator from the levels of ξ_{jt} .

We also assume a linear selection equation:

$$d_{jt} = \mathbf{1}(18 + w_{jt} - 5p_{jt} + \eta_{jt} > 0)$$

where $w_{jt} \sim$ i.i.d. $U[0, 1]$. Similarly, we assume linear prices:

$$p_{jt} = \gamma_0 + \sum_l x_{jlt} \gamma_l + (\tilde{z}_{j1t} + \tilde{z}_{j2t}) + \frac{\xi_{jt}}{2} + \nu_{jt}$$

where $\gamma = (1, \frac{1}{2}, 1, 1)'$, $\tilde{z}_{jkt} \sim$ i.i.d. $U[0, 1]$ for $k \in \{1, 2\}$, and $\nu_{jt} \sim \mathcal{N}(0, 0.41)$. Note that this

design always includes price endogeneity in demand due to the dependence of a product j 's price on its unobserved demand shock, ξ_{jt} .

We use Monte Carlo simulation to generate the market shares:

$$s_{jt} = \frac{1}{R} \sum_{r=1}^R \frac{d_{jt} \exp\left(\beta_0^r + \alpha^r p_{jt} + \sum_{k=1}^3 x_{jkt} \beta_k^r + \xi_{jt}\right)}{1 + \sum_{i=1}^J d_{it} \exp\left(\beta_0^r + \alpha^r p_{it} + \sum_{k=1}^3 x_{ikt} \beta_k^r + \xi_{it}\right)}$$

where $R = 50,000$. We conduct 100 replications of the sampling experiment. Each simulated dataset has 10,000 product-market combinations.

In scenario (a), we introduce selection-on-unobservables into the demand process by assuming: $\begin{bmatrix} \xi_{jt} \\ \eta_{jt} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4.56 & 3 \\ 3 & 4.56 \end{bmatrix}\right)$. The typical dataset has approximately 5,800 non-zero market shares ($\sum_j \sum_t d_{jt} \approx 5800$) and, hence, 42% of the observations involve a zero-valued market share.

In scenario (b), we allow for zero market shares but eliminate the selection-on-unobservables by assuming: $d_{jt} \sim \text{i.i.d. Bernoulli}(0.58)$. That is, we randomly select around 58 products out of $J = 100$ to match the number of non-zero market share products in scenario (a). In this data-generating process, d_{jt} does not depend on either η_{jt} or ξ_{jt} and hence zero shares are missing at random.

In scenario (c), we use the standard data-generating process assumed in the literature where $d_{jt} = 1, \forall j, t$. Since all products have strictly positive shares, it mechanically provides more information than scenarios (a) and (b). We therefore reduce the number of products in each market to $J = 58$ to replicate the 5,800 non-zero shares we observe in scenarios (a) and (b).

5.2 Results

We now summarize the results across the 100 replications of the data-generating processes described in section 5.1. For each replication we estimate 6 versions of the model, two of them are Random-Coefficients Logit and the rest are Homogeneous Logit models (i.e., with no random coefficients). We report two different Random-Coefficients Logit estimators here: (1) PDWGMM, and (2) Drop-Zero. We use $N_h = 2,000$ random draws to compute the expected shares for each of these random coefficients estimators.

Table 1 reports the average coefficient estimates across the 100 replications for our two random coefficients estimators. In scenario (a) of Table 1, the PDWGMM estimator recovers the mean parameters quite accurately while, as expected, the Drop-Zero estimator exhibits considerable selection bias. PDWGMM also recovers the standard deviation parameters quite accurately except for the standard deviation on the price coefficient, which is estimated imprecisely. In contrast, each of the standard deviation parameters is estimated imprecisely under Drop-Zero. Most striking, the Drop-

Zero estimator overestimates the price coefficient by around 150%. However, our results confirm that the Drop-Zero estimator fails to recover the true coefficient values when the incidence of zero shares is selected on observed aspects of demand. The reason we only find bias in the price coefficient is likely due to the fact that prices are uncorrelated with observed product attributes in the data-generating process. As we will see in the empirical case study in section 6 below, the Drop-Zero estimator can lead to substantively different (biased) estimates for all the model parameters.

Table 1: Monte Carlo Data Estimation Results

DGP	(a) With Selection-on-Unobservables		(b) Without Selection-on-Unobservables		(c) No Zero Shares	
Random-Coefficients Logit						
	PDWGMM	Drop 0	PDWGMM	Drop 0	PDWGMM	Drop 0
$\bar{\alpha}$ [-2]	-2.045 (0.112)	-5.019 (0.810)	-2.098 (0.350)	-2.116 (0.363)	-1.989 (0.325)	-2.053 (0.306)
$\bar{\beta}_1$ [1]	0.941 (0.093)	1.085 (0.340)	0.959 (0.114)	0.992 (0.139)	0.961 (0.127)	1.024 (0.159)
$\bar{\beta}_2$ [2]	1.957 (0.124)	2.332 (0.498)	1.991 (0.169)	2.085 (0.199)	1.968 (0.200)	2.126 (0.287)
$\bar{\beta}_3$ [2]	1.944 (0.125)	2.357 (0.495)	1.984 (0.166)	2.065 (0.231)	1.943 (0.175)	2.094 (0.285)
σ_α [1]	0.335 (0.613)	3.803 (5.693)	1.081 (0.854)	1.188 (0.994)	0.810 (0.796)	1.004 (0.910)
σ_{β_1} [1]	0.941 (0.215)	0.945 (0.820)	0.868 (0.489)	0.822 (0.534)	0.840 (0.553)	0.839 (0.568)
σ_{β_2} [2]	1.859 (0.132)	1.366 (0.888)	1.886 (0.344)	1.829 (0.447)	1.891 (0.365)	1.797 (0.483)
σ_{β_3} [2]	1.926 (0.112)	1.408 (0.848)	1.977 (0.335)	1.886 (0.509)	2.036 (0.445)	1.966 (0.504)
No. Conv.	99	97	97	97	99	97

Notes. Column names indicate how the zero shares are handled during the estimation. True parameter values are reported in square brackets. The reported estimates are the means of each parameter across the converged trials of 100 replications of the data generating process. Standard deviations across converged trials are reported in the parentheses. *No. Conv.* row is the number of converged estimates out of 100 Monte Carlo replications.

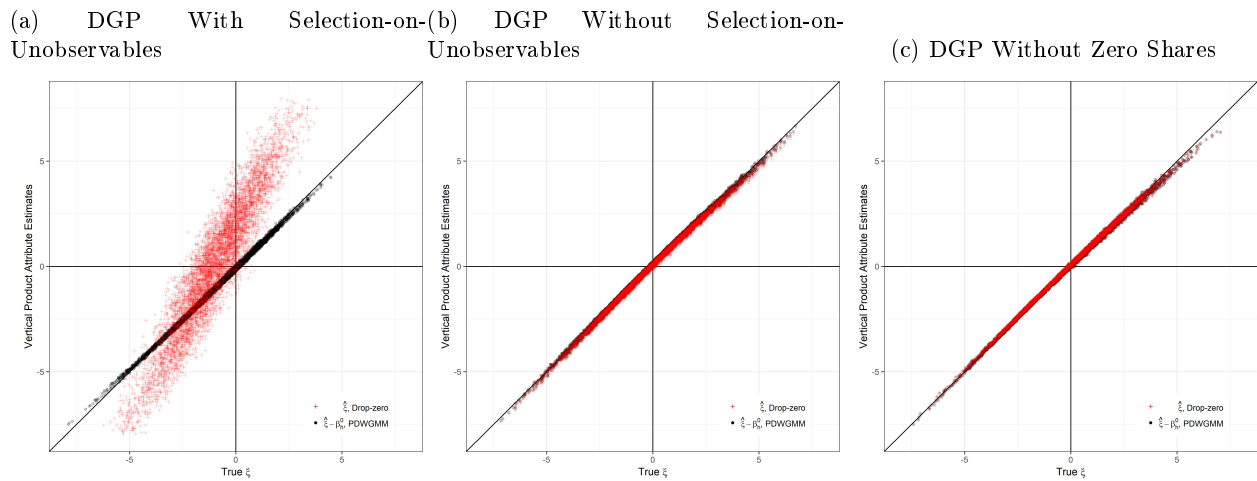
In scenario (b), Table 1 reports the estimation results when the observed zero market shares are missing at random. As expected, both estimators generate accurate and comparable estimates of the true model parameters. Importantly, PDWGMM does not appear to lose precision or accuracy relative to Drop-Zero (i.e., the standard Random-Coefficients Logit GMM estimator) when there is no selection on unobservables.

In scenario (c), Table 1 reports the estimation results when there are no observed zero market shares in the data. In this scenario, the PDWGMM estimator does not require the first two stages

since $\hat{\omega}_{i,j} = 1$ in stage 3 for all the observation pairs. The estimates in scenario (c) are very similar to those of scenario (b), with comparable results for both PDWGMM and Drop-Zero.

Appendix C reports the results from the Homogeneous Logit models which are computationally easier to implement. There too we find that the PDWGMM estimator provides relatively accurate and precise estimates of the mean taste parameters $(\bar{\alpha}, \bar{\beta})$, while the conventional methods return biased results in the presence of selection on unobservables in the incidence of zero market shares. For this particular data-generating process, we find that the Homogeneous and Random-Coefficients versions of the Drop-Zero estimator exhibit the same direction and comparable magnitudes of bias in the price coefficient estimates. We explore the direction of bias in more detail below, in section 5.3.⁹

Figure 5.1: Scatterplot of True ξ_j and $\hat{\xi}_j$ from Our PDWGMM Estimator and Drop-Zero Estimator, Monte Carlo Data



We next investigate the pattern of the relationship between the true ξ_j and estimated $\hat{\xi}_j$ for our PDWGMM estimator and Drop-Zero estimator, respectively. In our Monte Carlo exercises, we can net out the true value of the intercept term β_h^0 from the composite residual $\hat{\xi}_j$ to obtain the pure unobserved vertical product attribute term $(\hat{\xi}_j - \beta_h^0)$ for our PDWGMM estimator.¹⁰ Figures 5.1a to 5.1c each overlay two scatterplots using one of our Monte Carlo datasets for the observations with strictly positive shares, respectively for scenarios (a) to (c). The horizontal axes report the true ξ_j

⁹For the Homogeneous Logit model with an imputed share value, we also examined the sensitivity of the price coefficient estimates to the imputation value. Regardless of whether the incidence of zero market shares is selected on unobservables or not, the estimated price coefficient is very sensitive to the choice of imputation. For instance, in the presence of selection, the price coefficient estimates range from -3.605 (impute 10^{-5}) to -695.990 (impute 10^{-300}), and without the presence of selection, the price coefficient estimates range from -0.934 (impute 10^{-5}) to $+1.926$ (impute 10^{-300}). See Appendix B for our full sensitivity analysis. We also explore this sensitivity in our empirical case study in section 6.2.

¹⁰As explained in section 3.2.1, it would be infeasible to net out the intercept term from the composite residuals when implementing our PDWGMM estimator using real-world data.

values and the vertical axes report the estimated vertical product attribute values from different estimators.

The value of $(\hat{\xi}_j - \beta_h^0)$ from our PDWGMM estimator appears to estimate the true ξ_j very accurately and precisely. In Figure 5.1a, when selection on unobservables matters, the true ξ_j value and $(\hat{\xi}_j - \beta_h^0)$ from our PDWGMM estimation lie almost exactly on the 45° line through the origin. In contrast, $\hat{\xi}_j^{DZ}$ from the Drop-Zero estimator is much more dispersed, with the slope quite different from one. Second, the distribution of the true ξ_j for the observations with $s_j > 0$ is centered around the mean of -1.048 due to the selection on unobservables, and the distribution of $(\hat{\xi}_j^{PDWGMM} - \beta_h^0)$ from our PDWGMM estimator is centered around the mean of -1.119 , which is very close to the true value. In contrast, the distribution of $\hat{\xi}_j$ from the Drop-Zero estimator is mechanically centered around zero because of one of the sample moment condition being imposed during estimation, ignoring the selection on unobservables. In Figures 5.1b and 5.1c when selection on unobservables does not matter, both $(\hat{\xi}_j^{PDWGMM} - \beta_h^0)$ from our PDWGMM estimator and $\hat{\xi}_j^{DZ}$ from the Drop-Zero estimator lie almost exactly on the 45° line passing through the origin. As we demonstrate below, the direction of the bias in the Drop-Zero estimator depends in part on the strength of price endogeneity, and not just on selection. For other data-generating processes, $\mathbb{E}(\hat{\xi}_j - \beta_h^0) > 0$ and the Drop-Zero estimator can over-estimate the parameters.

5.3 Additional Monte Carlo Analysis on the Direction of Bias in Price-Coefficient Estimates

For the data-generating process explored in the previous section, selection-on-unobservables generated the same direction of bias and comparable magnitude of bias in the mean price coefficient for our Drop-Zero estimators. We now conduct additional sampling experiments to explore the direction of bias.

In practice, it is not possible to determine the direction of bias for the Drop-Zero estimator due to the simultaneous roles of price endogeneity and selection on unobservables. Under price endogeneity and selection, the sign of the bias in the price coefficient estimate for the Drop-Zero estimator depends on $\mathbb{E}[\xi_{jt} | \mathbf{z}_{jt}, d_{jt} = 1]$. If the non-selection propensity is decreasing in the unobserved demand shock, ξ_{jt} , then $\mathbb{E}[\xi_{jt} | \mathbf{z}_{jt}, d_{jt} = 1] > 0$. However, if the non-selection propensity is increasing in prices, then $\mathbb{E}[\xi_{jt} | \mathbf{z}_{jt}, d_{jt} = 1] < 0$, exacerbating the negative bias already induced when prices are positively correlated with unobserved demand shocks. The fact that there is no systematic direction of bias is important for applied work since it means that the popular Drop-Zero estimators are not only biased, they do not provide a systematic bound on the parameters of interest.

To explore this concern about an ambiguous direction of bias, we conduct two additional Monte Carlo experiments with selection on unobservables: (d) $\mathbb{E}[\xi_{jt} | \mathbf{z}_{jt}, d_{jt} = 1] < 0$ and (e) $\mathbb{E}[\xi_{jt} | \mathbf{z}_{jt}, d_{jt} = 1] > 0$. We focus on the Homogeneous Logit case to simplify the computational burden.

The data-generating process for scenario (d) is identical to the scenario (a) of section 5.1, except

that we set all the standard deviations of the random coefficients to zero: $\alpha^h = -2$, $\beta_0^h = 2$, $\beta_1^h = 1$, and $\beta_2^h = 2 \forall h$. Scenario (d) generates around 56% of nonzero share observations with $\mathbb{E}[\xi_{jt}|\mathbf{z}_{jt}, d_{jt} = 1] = -1.1$, which is negative.

The data generating processes for scenario (e) makes the following three modification on scenario (a) of section 5.1. First, $\alpha^h = -2$, $\beta_0^h = 2$, $\beta_1^h = 1$, and $\beta_2^h = 2 \forall h$, as in scenario (d). Second, we increase the variance in the non-selection outcomes as well as the covariance with the demand shock: $\begin{bmatrix} \xi_{jt} \\ \eta_{jt} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 6 \\ 6 & 26 \end{bmatrix}\right)$. Third, we weaken the dependence of consideration on prices: $d_{jt} = \mathbf{1}(w_{jt} - \frac{p_{jt}}{5} + \eta_{jt} > 0)$. Consequently, scenario (e) generates around 42% of nonzero share observations with $\mathbb{E}[\xi_{jt}|\mathbf{z}_{jt}, d_{jt} = 1] = 2.9$, which is positive.

We estimate Homogeneous Logit PDWGMM and Homogeneous Logit Drop-Zero estimators for scenarios (d) and (e), respectively. The results are reported in Table 2. The price coefficient estimate of the Drop-Zero estimator is biased upward (towards more negative) in scenario (d), whereas it is biased downward (towards more positive) in scenario (e). In contrast, our PDWGMM estimator recovers the target parameter accurately in both scenarios (d) and (e). In sum, the price coefficient estimates of the Drop-Zero estimator cannot be taken as either a lower or upper bound of the true parameter.

Table 2: Monte Carlo Data Estimation Results on the Direction of Bias of Price Coefficients

DGP	(d) $\mathbb{E}[\xi_{jt} \mathbf{z}_{jt}, d_{jt} = 1] < 0$	(e) $\mathbb{E}[\xi_{jt} \mathbf{z}_{jt}, d_{jt} = 1] > 0$	Homogeneous Logit	
Parameters	PDWGMM	Drop 0	PDWGMM	Drop 0
$\bar{\alpha}$ [-2]	-2.045 (0.059)	-4.635 (0.334)	-1.962 (0.144)	-1.628 (0.110)
$\bar{\beta}_1$ [1]	0.980 (0.091)	0.979 (0.172)	0.983 (0.205)	0.956 (0.208)
$\bar{\beta}_2$ [2]	1.974 (0.106)	1.978 (0.221)	1.997 (0.238)	1.953 (0.218)
$\bar{\beta}_3$ [2]	1.972 (0.104)	1.961 (0.197)	1.989 (0.242)	1.956 (0.236)

Notes. Column names indicate how the zero shares are handled during the estimation. True parameter values are reported in square brackets. The reported estimates are the means of each parameter across the 100 replications of the data generating process. Standard deviations across 100 trials are reported in the parentheses.

6 Application: Scanner Data with Many Zeros

6.1 Data and Evidence of Selection on Zero-Valued Market Shares

We conduct an empirical case study of the ice cream category. The data were provided by a midwestern chain that tracks daily prices and sales of each of the UPCs sold by store. A distinctive feature of these data is that they record shelf prices, unit costs, and feature advertisement information even when a UPC is not sold in a given store-day. The advantage of daily data is that we observe the exact price charged for each product, in contrast with most weekly-frequency datasets that report the average price for each product across each of the days in that week. We observe 25 stores during an 18-month period, generating 2.7 million UPC-store-day observations.

For our empirical case study, we focus on the non-premium segment of ice cream brands. We include Breyers, Deans, Edy’s and the Store Brand, which collectively account for 26.3% of the total category sales. Since 98.5% of sales for these brands come from the 48oz pack size, we only use 48oz packs. For estimation, we use 2 stores in the chain, retaining only the Thursdays since we observe very little within-week variation in prices. We use all the available data for each store between our sample range, January 2013 to June 2014.¹¹ Our estimation sample therefore includes 14,907 product-store observations, spanning 130 unique products and 131 unique markets (i.e., store-days).

To construct the potential market size, we start with each Thursday’s total unique customer count. Since it is unlikely that every store visit results in a trip to the non-premium section of the ice cream aisle, we scale the customer count by 1/7 of all trips.¹² Accordingly, we define each product’s market share on a store-day as: $s_j = \frac{\text{total packs of } j \text{ sold}}{\text{total potential ice cream purchases}}$. We define the outside good share, s_0 , as the share of the potential trips during which no non-premium ice cream packs were sold.

Zero shares are quite common in these retail data: only 30.7% of the store-day-product observations (e.g., 4,571 out of 14,907) generate at least one unit sale. The incidence of zero market shares appears to be more than random sampling error.¹³ Even if we collapse a given store’s product sales to a lower time frequency (e.g., weekly or monthly), we still observe many products with zero shares. Table 3 reports the incidence of zero shares when a store’s data are collapsed to the daily, weekly, monthly and tri-monthly frequencies, respectively. When we collapse the data to 90-day intervals, a typical store exhibits around 148,000 unique customer visits (i.e., 21,142 unique potential non-premium ice cream purchases)¹⁴ each making discrete choices from approximately 130 unique

¹¹One store opened before January 2013, the other store opened in June 2013.

¹²We use Nielsen Homescan data to estimate this scaling using the 271,016 unique household trips to this chain. Only 6,634 trips resulted in the purchase of a non-premium ice cream. We then set the potential market at 5-6 times this proportion of trips to account for consumers who considered by buying but ultimately did not, possibly due to prices or availability in other stores, generating a potential market size of 1/7 of the customer store visits to the chain in question.

¹³With a finite and discrete number of consumers each making a discrete choice decision, zero market shares could arise by pure chance.

¹⁴As in many retail databases, we observe each store’s total customer traffic counts.

non-premium ice cream products.¹⁵ Even, with this large consumer population, we still observe zero shares for 6.6% of the available non-premium ice cream products in the store-period. With this many underlying discrete choices, it is almost impossible for random utility alone to explain the incidence of zeros. Furthermore, we find that for 18.9% of the non-premium ice cream products exhibiting a zero share in at least one store-period, the minimum non-zero quantity sold is larger than 5 units. This type of discrete jump from zero to a large positive quantity also seems inconsistent with pure random utility explaining the zero shares. In sum, these patterns indicate systematic non-selection for certain products on the shelves.

Table 3: Average Number of Zero Shares of Ice Cream Per Store

Data Frequency	Total Number of packs sold	Total Unique Number of Customer Visits per Store	Number of Observations with Zero Sales	Average Number of Products Available Per Store
1 day	74	1,664	78	120
7 days	513	11,615	26	120
30 days	2,159	49,594	14	121
90 days	6,203	147,923	8	122

In addition to the brand name, we also observe an extensive set of product attribute dummy variables: Vanilla, Chocolate, Yogurt, Caramel, Mint, Gelato, and Fruit. For each product, we manually collected total calories per cup as a continuously-valued nutrient-related attribute.¹⁶¹⁷

To address the potential endogeneity of shelf prices, we use the unit cost variable and its square as price instruments, \tilde{z}_j . As in Chintagunta et al. (2003), we assume the wholesale prices and temporary discounts of each UPC are independent of a given store’s demand shocks.¹⁸ As in the Monte Carlo simulations, we use a linear price specification. To address the potential selection on consideration, we use UPC and brand-level feature advertisement variables as the consideration instruments \tilde{w}_j . Following Mehta et al. (2004), we assume these instruments affect consideration, but not demand. Intuitively, consumers do not derive consumption benefits from viewing a feature ad. However, the ad impression triggers their attention to the specific featured products and brands. Of course, these excluded promotional variables would also be valid if they affect stock-out propensities, another cause of non-consideration. In our daily data, it is possible that a promoted item stocks-out midway through the promotional period (typically several days), thereby generating zero sales on a given

¹⁵This large number of available product variants is approximately average for a CPG product category. Dubé (2019) finds that the average CPG category has about 400 available products, spanning 64 brands.

¹⁶To ensure all the continuous attributes are on a comparable scale for estimation, we use normalize the *calories per cup* information in the nutrition label by 1/1000.

¹⁷We collected the nutrient data from the following websites: www.walmart.com, www.calorieking.com, www.picknsave.com, www.breyers.com, www.target.com, www.edys.com, www.homemadebrand.com, www.kemps.com.

¹⁸Manufacturers do not target wholesale prices differentially across stores in a single chain.

day. Once again, an advantage of our non-parametric control for the consideration propensity is the robustness of our approach to different mechanisms.

Table 4 presents the key summary statistics of our estimation data. Further details about the data being used in this study can be found in section 2 of Bronnenberg et al. (Forthcoming).

Table 4: Summary Statistics of the Estimation Data

	Mean	Std. Dev	Min	Max
Price	3.835	0.461	1.895	4.990
Share	0.002	0.004	0	0
s_0	0.736	0.091	0.433	0.894
Cost/pack	4.164	0.508	0.000	4.530
Featured UPC	0.132	0.338	0	1
Featured Brand	0.190	0.392	0	1
Temporary Discount	0.036	0.186	0	1
Cal./(1000×cup)	0.265	0.043	0.18	0.38
Vanilla	0.198	0.398	0	1
Chocolate	0.195	0.396	0	1
Frozen Yogurt	0.132	0.338	0	1
Mint	0.084	0.278	0	1
Fruit	0.919	0.272	0	1
Breyers	0.205	0.403	0	1
Deans	0.228	0.420	0	1
Edys	0.532	0.499	0	1
Store Brand	0.035	0.184	0	1
Non-zero Observations		4571		
Sample Size		14907		

Notes. This table presents the summary statistics of the estimation data being used.

6.2 Estimation and Results

Table 5 presents the ice cream demand estimation results. We compare 6 alternative estimators and implementations. The first two columns labeled *Random Coefficient Logit* are the Random-Coefficients Logit model estimates with our PDWGMM selection correction and Drop-Zero estimators, respectively.¹⁹ The next four columns labeled *Homogeneous Logit* are Homogeneous Logit model estimates of our selection-correction (PDWGMM), classical Heckman correction, Drop-Zero, and Impute-Zero estimators, respectively. Each estimator uses the same set of instruments for a fair comparison.

¹⁹The Random-Coefficient Logit models are estimated using $N_h = 2,000$ simulation draws, and we did not include the taste heterogeneity parameters in the brand coefficients. Our PDWGMM estimator used the bandwidth following Silverman’s rule throughout the estimation. See Appendix A.5 for details on the choice of bandwidth.

Table 5: Model Parameter Estimation Results

	Random-Coefficients Logit				Homogeneous Logit				
	PDWGMM		Drop 0		PDWGMM	Heckman	Drop 0	Impute 10^{-12}	Drop 0 OLS
	$\bar{\theta}$	$\sqrt{\Sigma}$	$\bar{\theta}$	$\sqrt{\Sigma}$					
Mean Elast.	-2.512		-1.489		-2.648	-1.939	-1.394	-17.625	-1.290
Price	-0.689 (0.123)	0 (0.002)	-0.411 (0.069)	0.002 (0.001)	-0.716 (0.147)	-0.524 (0.057)	-0.377 (0.061)	-4.767 (0.702)	-0.349 (0.022)
Cal./(1000×cup)	-0.166 (0.224)	0 (0.157)	0.024 (0.253)	0.001 (0.012)	-0.235 (0.191)	-0.116 (0.214)	-0.141 (0.205)	-0.953 (2.312)	-0.131 (0.204)
Vanilla	0.173 (0.027)	0.002 (0.012)	0.175 (0.028)	0 (0.002)	0.156 (0.020)	0.159 (0.020)	0.146 (0.020)	2.651 (0.248)	0.145 (0.020)
Chocolate	-1.088 (1.098)	1.760 (1.156)	-3.128 (1.874)	3.337 (1.688)	0.044 (0.019)	0.054 (0.020)	0.044 (0.020)	0.741 (0.234)	0.044 (0.020)
Yogurt	-0.140 (1.408)	0.101 (0.999)	-0.129 (0.104)	0 (0.187)	-0.122 (0.025)	-0.132 (0.028)	-0.129 (0.028)	-2.476 (0.243)	-0.128 (0.028)
Mint	-2.829 (2.275)	2.849 (1.664)	-1.664 (1.269)	2.174 (1.239)	0.018 (0.029)	0.008 (0.027)	0.007 (0.027)	0.806 (0.356)	0.007 (0.027)
Fruit	-0.121 (0.034)	0 (0.004)	-0.085 (0.033)	0 (0.001)	-0.106 (0.032)	-0.071 (0.031)	-0.067 (0.03)	-0.124 (0.320)	-0.066 (0.030)
Breyers	0.237 (0.136)	0 (0.014)	0.158 (0.068)	0 (0.052)	0.167 (0.159)	0.264 (0.062)	0.159 (0.063)	-1.720 (0.796)	0.138 (0.046)
Deans	0.230 (0.234)	0.159 (0.398)	0.082 (0.109)	0.007 (0.222)	0.234 (0.165)	0.301 (0.064)	0.100 (0.069)	-4.772 (1.002)	0.074 (0.045)
Edys	0.243 (0.144)	0 (0.099)	0.105 (0.070)	0 (0.003)	0.172 (0.158)	0.249 (0.059)	0.078 (0.063)	-3.991 (0.884)	0.055 (0.043)
No. Conv.	482		500		500	500	500	500	500

Notes. The sub-columns $\bar{\theta}$ are the mean parameter estimates, $\sqrt{\Sigma}$ are the standard-deviation parameter estimates. Bootstrap standard errors are in parentheses based on $B = 500$ draws. *Mean Elast.* row indicates the mean price elasticities across all the products available in our sample, using the usual formula. *No. Conv.* row indicates the number of converged bootstrap trials out of 500 bootstrap iterations.

As in the Monte Carlo analysis in section 5, the imputed share correction is found to be highly unstable and results vary dramatically with the choice of imputation value. For the Homogeneous Logit, Appendix B indicates that the price coefficient estimates vary from -0.917 to -143.362 as we vary the imputed share value from 10^{-4} to 10^{-300} . The lack of robustness of this estimator suggests it does not provide a reliable estimate of demand and we do not discuss it further in this section.

The PDWGMM estimator generates a much smaller mean price coefficient than the Drop-Zero estimator, in this case a difference of 0.278 . We can test the significance of this difference formally using a two-sample bootstrap t -test of the null hypothesis that the price coefficients are the same:^{20,21}

$$T_B = \frac{|\hat{\alpha}^{PDWGMM} - \hat{\alpha}^{drop0}|}{\sqrt{Var(\hat{\alpha}^{PDWGMM}) + Var(\hat{\alpha}^{drop0}) - 2Cov(\hat{\alpha}^{PDWGMM}, \hat{\alpha}^{drop0})}}.$$

We compute $T_B = 2.26$ (p-value of 0.023), thereby rejecting the hypothesis of equal mean price coefficients at the 5% significance level. At the 5% significance level, we can rule out differences smaller than 0.038 .

Our two estimators also generate very different own-price elasticities, with a mean price elasticity of -1.489 for the Drop-Zero estimator and a mean price elasticity of -2.512 for PDWGMM. We reject the hypothesis of equal mean elasticities at the 5% significance level, with $T_B = 2.34$ (p-value of 0.019).

We find similar results for the computationally lighter Homogeneous Logit model of demand in the final four columns. Interestingly, we generally find similar estimates of mean tastes with and without random-coefficients. However, as before, the price coefficient estimate of PDWGMM is smaller (more negative) than that of the Drop-Zero estimator. As before, we reject the hypothesis of equal mean price coefficients for the PDWGMM and Drop-Zero estimators at the 5% significance level, with $T_B = 2.54$ (p-value of 0.011). Even the classical Heckman two-step estimator fails to resolve the selection bias, generating a similar price coefficient estimates as the Drop-Zero estimator.

In sum, our PDWGMM estimator generates significantly different price coefficients and price elasticities than the usual Drop-Zero estimator. These differences are consistent with our Monte Carlo evidence in in section 5. In the next section, we use the observed retail margins in our data to construct a formal test between the PDWGMM and Drop-Zero estimators.

6.3 Assessing Predictive Model Fit Using Retail Markups

We now use the retail cost information in our data to construct a test between the PDWGMM estimator and the Drop-Zero estimator based on implied mark-ups under various retail pricing

²⁰The covariance term $Cov(\hat{\alpha}^{PDWGMM}, \hat{\alpha}^{drop0})$ requires pre-determining the bootstrap resample order \mathbf{n}_b and the household-specific taste shocks ν^h to ensure the two estimators use the same bootstrap samples.

²¹We set the max iteration count to 200. For PDWGMM, 482 of the 500 bootstrap trials converged and all 500 bootstrap trials of the Drop-Zero estimator converged. For the test, we dropped the 18 bootstrap trials for which our PDWGMM estimator failed to converge within 200 iterations.

conduct assumptions. We assume the retailer sets prices to maximize its static profits in the ice cream category, where $\Pi(\mathbf{p}; \Theta) = (\mathbf{p} - \mathbf{c})' Q(\mathbf{p}; \Theta)$ denotes the total category profits and $Q(\mathbf{p}; \Theta) = M\boldsymbol{\pi}(\mathbf{p}, \tilde{\mathbf{X}}, \boldsymbol{\xi}; \Theta)$ denotes the predicted vector of quantities sold for each UPC. Let \mathbf{m} denote the vector of observed margins, with element $m_j = p_j - c_j$ for product j . The first-order necessary conditions for optimal pricing are:

$$\boldsymbol{\pi}(\mathbf{p}^*; \Theta) + [\nabla_{\mathbf{p}} \boldsymbol{\pi}(\mathbf{p}^*; \Theta) \circ \boldsymbol{\Xi}] (\mathbf{p}^* - \mathbf{c}) = \mathbf{0}, \quad (6.1)$$

where $\nabla_{\mathbf{p}} \boldsymbol{\pi}(\mathbf{p}; \Theta)$ is the gradient of the demand system with (i, j) element:

$$\frac{\partial \pi_j}{\partial p_i} = \begin{cases} \int (\alpha + \nu_\alpha) \pi_j^h(\mathbf{p}) (1 - \pi_j^h(\mathbf{p})) dF_{\boldsymbol{\nu}}, & \text{if } i = j \\ - \int (\alpha + \nu_\alpha) \pi_i^h(\mathbf{p}) \pi_j^h(\mathbf{p}) dF_{\boldsymbol{\nu}} & \text{if } i \neq j \end{cases}$$

and $\boldsymbol{\Xi}$ is a matrix of zeros and ones reflecting various forms of retail conduct. The optimal contribution margin can then be written as:

$$\boldsymbol{\mu}(\Theta) = -[\nabla_{\mathbf{p}} \boldsymbol{\pi}(\mathbf{p}^*; \Theta) \circ \boldsymbol{\Xi}]^{-1} \boldsymbol{\pi}(\mathbf{p}^*; \Theta). \quad (6.2)$$

Since the retailer's contribution margins are observed, we can construct a minimum distance estimator of the margins based on the criterion:

$$q(\boldsymbol{\mu}) = (\mathbf{m} - \boldsymbol{\mu}^{Model})' (\mathbf{m} - \boldsymbol{\mu}^{Model}). \quad (6.3)$$

Of particular interest are the margin estimates $\boldsymbol{\mu}(\Theta^{PDWGMM})$ and $\boldsymbol{\mu}(\Theta^{Drop0})$ using equation (6.2). We can then test the relative fit of the different demand estimators based on their ability to fit the observed contribution margins: $q(\Theta^{PDWGMM})$ and $q(\Theta^{Drop0})$. We select the model with a smaller criterion value.

For robustness, we conduct three separate model selection tests, each under a different assumed form of retail pricing conduct:

(a) *Joint Monopoly* whereby the retailer sets all the prices jointly to maximize total profits ($\boldsymbol{\Xi}$ is a matrix of ones for the same market);

(b) *Brand-level Pricing* whereby the retailer maximizes the joint profits of products sold under a common brand (the entries of $\boldsymbol{\Xi}$ is 1 if products share a common brand name within the same market, and 0 otherwise); and

(c) *UPC-level Pricing* whereby the retailer optimizes each of the products' prices in the category independently of one-another ($\boldsymbol{\Xi}$ is an identity matrix).

Table 6 presents the model-fit comparison measures for the random-coefficients specifications of PDWGMM and Drop-Zero under each of the three conduct assumptions. The PDWGMM estimator

of the demand parameter generates a considerably better fit of the retail contribution margins under all three forms of conduct. For each conduct assumption, the PDWGMM margins are “closer” to the true margins than the values from the Drop-Zero estimator. Figure 6.1 illustrates the difference between predicted margins and observed margins, respectively under the three retail pricing conduct assumption. The vertical line at 0 indicates that the difference between the predicted margin and the observed margin is zero, implying that the model predicts the observed margin perfectly. As a histogram is more concentrated around the vertical line at 0, the model has a better fit. Although both our PDWGMM and Drop-Zero estimators overestimates the margins, it turns out that our PDWGMM estimator performs better than the Drop-Zero estimator under all three conduct assumptions. The figures also indicate that the UPC-level pricing conduct assumption explains the observed margin the best among the three conduct assumptions we examined.

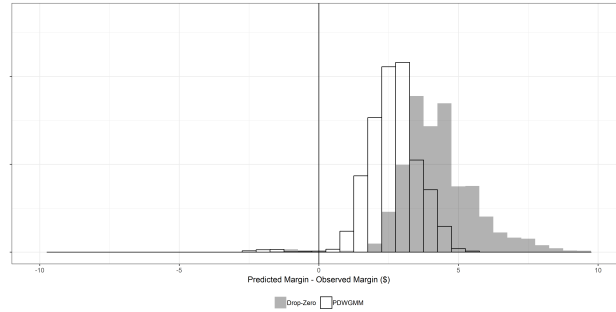
Table 6: A Comparison of Predictive Model Fit on Markups

	q^{PDWGMM}	$q^{Drop 0}$
(a) Joint Monopoly	36,018 (297)	92,021 (824)
(b) Brand-level Pricing	29,210 (257)	69,404 (625)
(c) UPC-level Competitive Pricing	20,383 (172)	43,049 (301)
Sample Size	4,571	

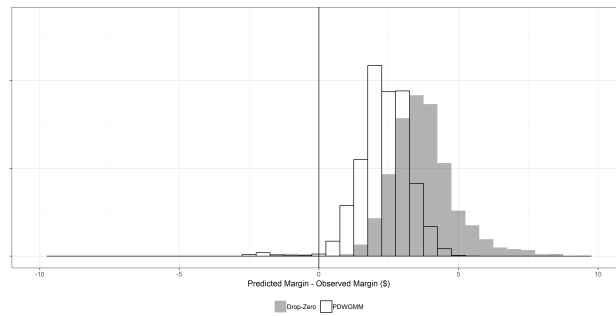
Notes. This table compares the measure of demand model prediction on markups (6.3) for three different conducts. Bootstrapped standard errors are reported in parentheses.

Figure 6.1: Histogram of the Differences Between Predicted Margins and Observed Margins

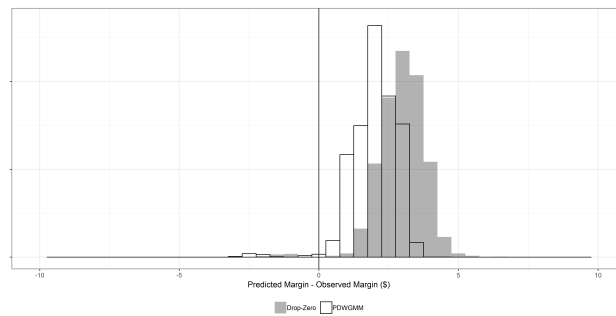
(a) Margin Difference Under Joint Monopoly Conduct Assumption



(b) Margin Difference Under Brand-Level Pricing Conduct Assumption



(c) Margin Difference Under UPC-Level Competitive Pricing Conduct Assumption



7 Extension: Share-Equation Equivalence of a CES Demand System and a Random-Coefficients Logit Demand System

In this section, we show that our proposed approach extends to the case of the CES model of demand which has been used extensively in the trade literature (e.g., Spence 1976; Dixit and Stiglitz 1977; Krugman 1980; Feenstra 1994; Handbury 2013; Handbury and Weinstein 2014; Bronnenberg 2015).

We show the equivalence between the Random-Coefficients Logit demand system derived in section 2.1 and the representative consumer demand system under CES preferences. It then follows that our PDWGMM estimator can be applied to the CES demand model to accommodate observations of zero aggregate sales for a product.

We modify the demand model derived in section 2.1. We relax the discrete-choice assumption and allow $q_j \in \mathbb{R}_+$. We also assume that consumers make purchase decisions to maximize the following CES preferences utility function

$$u(\{q_j, \tilde{\mathbf{x}}_j, \xi_j\}_{j \in \mathcal{J}}) := \left(\sum_{j \in \mathcal{J}} \{\chi(\tilde{\mathbf{x}}_j, \xi_j)\}^{\frac{1}{\sigma}} q_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad (7.1)$$

where $\chi(\tilde{\mathbf{x}}_j, \xi_j) \geq 0$ is a quality kernel, $\tilde{\mathbf{x}}_j$ is the vector of exogenous product characteristics that is observable to the econometrician, and ξ_j is a vertical product characteristics that is unobserved to the researcher as before.

The representative consumer's budget-constrained utility maximization problem is:

$$\max_{\{q_j\}_{j \in \mathcal{J}}} \left(\sum_{j \in \mathcal{J}} \{\chi(\tilde{\mathbf{x}}_j, \xi_j)\}^{\frac{1}{\sigma}} q_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad s.t. \quad \sum_{j \in \mathcal{J}} p_j q_j = y.$$

The corresponding Marshallian demand system is:

$$q_j = y \left\{ \frac{\chi(\tilde{\mathbf{x}}_j, \xi_j) p_j^{-\sigma}}{\sum_{k \in \mathcal{J}} \chi(\tilde{\mathbf{x}}_k, \xi_k) p_k^{1-\sigma}} \right\} \quad \forall j \in \mathcal{J}, \quad (7.2)$$

which can be re-written in the following market share format:

$$\pi_j \equiv \frac{q_j}{\sum_{k \in \mathcal{J}} q_k} = \frac{\chi(\tilde{\mathbf{x}}_j, \xi_j) p_j^{-\sigma}}{\sum_{k \in \mathcal{J}} \chi(\tilde{\mathbf{x}}_k, \xi_k) p_k^{-\sigma}}. \quad (7.3)$$

Let $h = 1, \dots, H$ index households, each with potentially heterogeneous quality kernel $\chi^h(\cdot)$ and elasticity of substitution σ^h . The individual quantity share expressions of the CES demand system (7.3) become:

$$\begin{aligned} \pi_j^h &= \frac{\chi^h(\tilde{\mathbf{x}}_j, \xi_j) \exp(-\sigma^h \ln p_j)}{\sum_{k=0}^J \chi^h(\tilde{\mathbf{x}}_k, \xi_k) \exp(-\sigma^h \ln p_k)} \\ &= \frac{\exp(-\sigma^h \ln p_j + \tilde{\mathbf{x}}_j' \boldsymbol{\beta}^h + \xi_j)}{\sum_{k=0}^J \exp(-\sigma^h \ln p_k + \tilde{\mathbf{x}}_k' \boldsymbol{\beta}^h + \xi_k)}, \end{aligned} \quad (7.4)$$

where the second equality follows by specifying $\chi^h(\tilde{\mathbf{x}}_j, \xi_j) = \exp(\tilde{\mathbf{x}}_j' \boldsymbol{\beta}^h + \xi_j)$. The predicted ag-

gregate market share equation is obtained by aggregating the individual predicted quantity share equations over the households h .

The CES market share is almost identical to the Random-Coefficients Logit demand except for some subtle differences. First, prices enter the demand system in logarithmic form. Second, the budget-constraint $\sum_{j \in \mathcal{J}} p_j q_j = y$ must be imposed to infer the vector of quantities demanded, \mathbf{q} . In contrast, for the discrete choice demand model, the fixed market size constraint $\sum_{j \in \mathcal{J}} q_j = M$ is used to infer the vector of quantities demanded, \mathbf{q} . Finally, the near mathematical equivalence only holds for the stylized case of Type I extreme value random utility errors in the discrete choice formulation.

8 Conclusion

We proposed a new method to handle differentiated products demand estimation in the presence of products capturing zero market share. We incorporate the role of selection on unobservables into the model to correct for the potential bias when observations with zero market shares are dropped from the demand estimation. We also propose a new, PDWGMM estimator that accommodates this selection under a fully non-parametric specification during the selection stage. An empirical case study shows that the PDWGMM estimator generates more price-elastic estimates of demand and a better fit of retail contribution margins.

In practice, many consumer sales datasets exhibit a high incidence of zero-valued shares. Our findings suggest that typical, ad hoc fixes will generate inconsistent demand estimates which would likely have adverse implications for policy experiments using the estimated demand system. In sum, we document a different form of endogeneity bias, selection on unobservables in product consideration, that can bias demand estimates in addition to the biases from endogenous marketing variables described in the extant literature.

One potential extension of the proposed framework would be to study the potentially self-selected manner in which assortments are set across markets, such as cities or even stores. Determining appropriate exclusion restrictions to determine profit instruments for the selection stage would be an interesting direction for future research pooling demand estimation across markets with different assortments.

Numerous mechanisms could explain the selection on unobservables associated with zero-valued shares, including consumer consideration, stock-outs and promotional conditions. An interesting direction for future research would consist of supplementing the point-of-sale data with additional moments on factors such as inventory and/or consumer search and consideration to allow for a richer model of how products ultimately enter consumers's choice sets. Moreover, data from these alternative mechanisms could provide additional information to help with the identification of the deep, consumer preference parameters.

References

- ABALUCK, J. AND A. ADAMS (2018): “What do consumers consider before they choose? Identification from asymmetric demand responses,” *Working Paper*.
- ADAO, R., A. COSTINOT, AND D. DONALDSON (2017): “Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade,” *American Economic Review*, 107, 633–689.
- AHN, H. AND J. L. POWELL (1993): “Semiparametric estimation of censored selection models with a nonparametric selection mechanism,” *Journal of Econometrics*, 58, 3–29.
- ARADILLAS-LOPEZ, A. (2012): “Pairwise-difference estimation of incomplete information games,” *Journal of Econometrics*, 168, 1.
- ARADILLAS-LOPEZ, A., B. E. HONORÉ, AND J. L. POWELL (2007): “Pairwise difference estimation with nonparametric control variables,” *International Economic Review*, 48, 1119–1158.
- BERRY, S. (1994): “Estimating discrete-choice models of product differentiation,” *RAND Journal of Economics*, 25, 242–262.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile prices in market equilibrium,” *Econometrica*, 63, 841–890.
- BESANKO, D., S. GUPTA, AND D. JAIN (1998): “Logit demand estimation under competitive pricing behavior: An equilibrium framework,” *Management Science*, 44, 1533–1547.
- BLUNDELL, R. W. AND J. L. POWELL (2004): “Endogeneity in semiparametric binary response models,” *Review of Economic Studies*, 71, 655–679.
- BRONNENBERG, B. J. (2015): “The provision of convenience and variety by the market,” *RAND Journal of Economics*, 46, 480–498.
- BRONNENBERG, B. J., J.-P. DUBÉ, AND R. E. SANDERS (Forthcoming): “Consumer misinformation and the brand premium: A private level blind taste test,” *Marketing Science*.
- BRONNENBERG, B. J., J. B. KIM, AND C. F. MELA (forthcoming): “Zooming in on Choice: How Do Consumers Search for Cameras Online?” *Marketing Science*.
- BRUNO, H. A. AND N. J. VILCASSIM (2008): “Structural demand estimation with varying product availability,” *Marketing Science*, 27, 1126–1131.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and confidence regions for parameter sets in econometric models,” *Econometrica*, 75, 1243–1284.

- CHINTAGUNTA, P., J.-P. DUBÉ, AND V. SINGH (2003): “Balancing Profitability and Consumer Welfare in a Supermarket Chain,” *Quantitative Marketing and Economics*, 1, 111–147.
- CHINTAGUNTA, P., T. ERDEM, AND P. E. R. ABD MICHEL WEDEL (2006): “Structural Modeling in Marketing: Review and Assessment,” *Marketing Science*, 25, 604–616.
- CILIBERTO, F. AND E. TAMER (2009): “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, 77, 1791–1828.
- CONLON, C. T. AND J. H. MORTIMER (2013): “Demand Estimation under Incomplete Product Availability,” *AMERICAN ECONOMIC JOURNAL: MICROECONOMICS*, 5, 1–30.
- COOPER, L. G. AND M. NAKANISHI (1988): *Market Share Analysis*, Kluwer Academic Publishers.
- DIXIT, A. K. AND J. E. STIGLITZ (1977): “Monopolistic competition and optimum product diversity,” *American Economic Review*, 67, 297–308.
- DUBÉ, J.-P. (2019): *The Handbook of the Economics of Marketing*, Noth Holland, chap. Microeconomic Models of Demand, 1–68.
- DUBÉ, J.-P., J. T. FOX, AND C.-L. SU (2012): “Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation,” *Econometrica*, 80, 2231–2267.
- FEENSTRA, R. C. (1994): “New product varieties and the measurement of international prices,” *American Economic Review*, 84, 157–177.
- GANDHI, A. AND J.-F. HOUDE (2019): “Measuring substitution patterns in differentiated products industries,” *Working Paper*.
- GANDHI, A., Z. LU, AND X. SHI (2019): “Estimating Demand for Differentiated Products with Zeroes in Market Share Data,” *Working Paper*.
- GUPTA, S., P. CHINTAGUNTA, A. KAUL, AND D. R. WITTINK (1996): “Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data,” *Journal of Marketing Research*, 33, 383–398.
- HANDBURY, J. (2013): “Are poor cities cheap for everyone?: Non-homotheticity and the cost of living across U.S. cities,” *Working Paper*.
- HANDBURY, J. AND D. E. WEINSTEIN (2014): “Goods prices and availability in cities,” *Review of Economic Studies*, 82, 258–296.
- HAUSER, J. R. (1978): “Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach,” *Operations Research*, 26, 406–421.

- HAUSER, J. R. AND B. WERNERFELT (1990): “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16, 393–408.
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating trade flows: Trading partners and trading volumes,” *Quarterly Journal of Economics*, 123, 441–487.
- HONKA, E. (2014): “Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry,” *Rand Journal of Economics*, 45, 847–884.
- HONORÉ, B. E. AND J. L. POWELL (2005): *Pairwise difference estimators for nonlinear models*, Cambridge University Press, chap. 22, 520–553, Identification and Inference for Econometric Models.
- IMBENS, G. W. AND C. F. MANSKI (2004): “Confidence intervals for partially identified parameters,” *Econometrica*, 72, 1845–1857.
- JOO, M., D. K. GAURI, AND K. C. WILBUR (Forthcoming): “Temporal distance and price responsiveness: Empirical investigation of the cruise industry,” *Management Science*.
- KIM, J. B., P. ALBUQUERQUE, AND B. J. BRONNENBERG (2010): “Online Demand Under Limited Consumer Search,” *Marketing Science*, 29, 1001–1023.
- (2017): “The Probit Choice Model under Sequential Search with an Application to Online Retailing,” *Management Science*, 63, 3911–3929.
- KRUGMAN, P. (1980): “Scale economies, product differentiation, and the pattern of trade,” *American Economic Review*, 70, 950–959.
- LI, Q. AND J. S. RACINE (2008): “Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data,” *Journal of Business and Economic Statistics*, 26, 423–434.
- MANSKI, C. F. AND E. TAMER (2002): “Inference on regressions with interval data on a regressor or outcome,” *Econometrica*, 70, 519–546.
- MATZKIN, R. L. (1992): “Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models,” *Econometrica*, 60, 239–270.
- MEHTA, N., S. RAJIV, AND K. SRINIVASAN (2003): “Price uncertainty and consumer search: A structural model of consideration set formation,” *Marketing Science*, 22, 58–84.
- (2004): “Role of Forgetting in Memory-Based Choice Decisions: A Structural Model,” *Quantitative Marketing and Economics*, 2, 107–140.

- MOORTHY, S., B. T. RATCHFORD, AND D. TALUKDAR (1997): "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23, 263–277.
- NEVO, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.
- (2011): "Empirical models of consumer behavior," *Annual Review of Economics*, 3, 51–75.
- PETRIN, A. AND K. E. TRAIN (2010): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47, 3–13.
- POWELL, J. L. (2001): *Semiparametric estimation of censored selection models*, Cambridge: Cambridge University Press, chap. 6, 165–196.
- RATCHFORD, B. T. (1980): "The value of information for selected appliances," *Journal of Consumer Research*, 17, 14–25.
- SHUGAN, S. (1980): "The cost of thinking," *Journal of Consumer Research*, 17, 99–111.
- SPENCE, M. (1976): "Product selection, fixed costs, and monopolistic competition," *Review of Economic Studies*, 43, 217–235.
- TENN, S. (2006): "Avoiding aggregation bias in demand estimation: A multivariate promotional disaggregation approach," *Quantitative Marketing and Economics*, 383–405.

Appendix

A Implementation Details

A.1 Construction of Instruments

For the GMM problem, we use the unit cost and temporary discount flag and squared unit cost, $\tilde{\mathbf{Z}}$, to instrument for prices. In addition, we use Gandhi and Houde (2019)'s *product differentiation* instruments to construct auxiliary moments based on functions of the product attributes, $\tilde{\mathbf{X}}$ and $\hat{\mathbf{p}}$, where $\hat{\mathbf{p}}$ is the vector of predicted prices from the pricing equation (4.2), for the identification of the standard deviations of the random coefficients, Σ (see also Joo et al., Forthcoming, for an application). For continuous product attributes and predicted prices, two products are regarded as *neighbors* when they are within 1 standard deviation of the corresponding attribute variable. For binary product attributes, two products are regarded as *neighbors* when they have the same value. For each SKU-level alternative, the differentiation variable corresponding to an attribute is defined by counting the number of within-market neighbors. The instrument matrix used for our PDWGMM estimator is then constructed as follows: $\mathbf{Z} := \left[\tilde{\mathbf{X}}' \tilde{\mathbf{Z}}' \hat{\mathbf{p}}_{\text{diff}}' \mathbf{X}_{\text{diff}}' \right]$, where the subscript *diff* denotes the product differentiation instrument. We then take pairwise-differences to form $\Delta \mathbf{Z}$. Finally, we define the GMM weight matrix, $\Phi := ((\Delta \mathbf{Z})' (\Delta \mathbf{Z}))^{-1}$.

For the drop-zero (BLP) estimator, we use the standard instrument matrix (i.e., undifferenced): $\mathbf{Z} := \left[\tilde{\mathbf{X}}' \tilde{\mathbf{Z}}' \hat{\mathbf{p}}_{\text{diff}}' \mathbf{X}_{\text{diff}}' \right]$. We then define the corresponding GMM weight matrix: $\Phi := (\mathbf{Z}' \mathbf{Z})^{-1}$.

A.2 MPEC Formulation for the Random-Coefficients Logit PDWGMM Estimator

We modify the MPEC formulation given in Dubé et al. (2012) to our selection-correction PDWGMM estimator of $\Theta := (\bar{\theta}, \Sigma)$. Let $\hat{\omega}_{i,j} := \frac{1}{h_n} \kappa \left(\frac{\hat{\mu}_i - \hat{\mu}_j}{h_n} \right)$. Let $(\sqrt{\Omega})$ be $\binom{n}{2} \times p$ matrix such that each column is $\sqrt{\hat{\omega}}$. Let $\mathbf{Q} := (\sqrt{\Omega}) \circ (\Delta \mathbf{Z})$ where \circ denotes the Hadamard product. We can then write $\mathbf{Q}' (\Delta \xi) = (\Delta \mathbf{Z})' (\sqrt{\hat{\omega}} \circ (\Delta \xi))$. We can now write the MPEC estimator (4.7) in matrix form:

$$\begin{aligned} & \min_{(\Theta, \xi, \mathbf{g})} \mathbf{g}' \Phi \mathbf{g} \\ \text{s.t. } & \mathbf{0} = \pi(\xi; \Theta) - \mathbf{s} \\ & \mathbf{0} = \mathbf{Q}' (\Delta \xi) - \mathbf{g}, \end{aligned} \tag{A.1}$$

where $\pi(\cdot)$ is the system of predicted market shares and Φ is the $p \times p$ GMM weighting matrix.

The Lagrangian is:

$$\mathcal{L}(\Theta, \xi, \mathbf{g}, \lambda_s, \lambda_g) = \mathbf{g}'\Phi\mathbf{g} + \lambda_s'(\pi(\xi; \theta) - \mathbf{s}) + \lambda_g'((\Delta\mathbf{Z})'(\Delta\xi_{\sqrt{\hat{\omega}}}) - \mathbf{g}).$$

The gradient of the objective function is:

$$\nabla \mathbf{g}'\Phi\mathbf{g} = \begin{pmatrix} \mathbf{0}' & \mathbf{0}' & 2\mathbf{g}'\Phi' \end{pmatrix}.$$

The constraint Jacobian is:

$$\begin{pmatrix} \frac{\partial \pi}{\partial \Theta} & \frac{\partial \pi}{\partial \xi} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{Q}'(\Delta\xi)}{\partial \xi} & -\mathbf{I} \end{pmatrix}.$$

$\frac{\partial \pi}{\partial \Theta}$ and $\frac{\partial \pi}{\partial \xi}$ part are identical to what is described in the Appendix of Dubé et al. (2012).

To derive $\frac{\partial \mathbf{Q}'(\Delta\xi)}{\partial \xi}$, first let $\Psi = \left(\frac{\partial \mathbf{Q}'(\Delta\xi)}{\partial \xi}\right)'$. Recall $\mathbf{Q} := (\sqrt{\hat{\Omega}}) \circ (\Delta\mathbf{Z})$ and the structure of $\Delta\xi$:

$$\Delta\xi := \begin{pmatrix} \xi_1 - \xi_2 \\ \xi_1 - \xi_3 \\ \vdots \\ \xi_1 - \xi_n \\ \xi_2 - \xi_3 \\ \xi_2 - \xi_4 \\ \vdots \\ \xi_2 - \xi_n \\ \vdots \\ \xi_{n-1} - \xi_n \end{pmatrix}.$$

It can be shown that the k th row of Ψ is given by:

$$\psi'_k = \mathbf{z}'_k \left(\sum_{l=1}^n \sqrt{\hat{\omega}_{k,l}} \right) - \sum_{l=1}^n \left(\sqrt{\hat{\omega}_{k,l}} \mathbf{z}'_l \right),$$

where \mathbf{z}'_k is k th row of the instrument matrix \mathbf{Z} .

A.3 Bootstrap Algorithm for the Random-Coefficients Logit PDWGMM Estimator

Suppose, first, that we are estimating the original BLP without selection-correction. Bootstrapping would require first inverting the whole vector ξ out from the original set of market share constraints

$$\mathbf{0} = \pi(\xi; \Theta) - \mathbf{s}$$

and then resample ξ_j with the corresponding row of \mathbf{Z} with replacement. This would also require the weighting matrix Φ to change with the bootstrap resample scheme. Keeping this intuition in mind, we describe the bootstrapping scheme below step-by-step:

1. Randomly draw n integers out of $(1, 2, \dots, n)$. Denote this vector of resample index integers by \mathbf{n}_b . Throughout, the subscript b denotes the resampled matrix/vector according to this resample scheme \mathbf{n}_b .
2. Take $(\mathbf{Z}_b, \mathbf{p}_b)$ from (\mathbf{Z}, \mathbf{p}) to estimate $\hat{\boldsymbol{\theta}}_b$ in (4.1) of section 4.1.
3. Take $\hat{\boldsymbol{\nu}}_b$ as the residual from the regression, to form the \mathbf{W}_b matrix corresponding to the matrix \mathbf{W} in section 4.1.
4. Find the vector $\hat{\boldsymbol{\mu}}_b$ in (4.4) where each element of $\hat{\boldsymbol{\mu}}_b$ is $\hat{\mu}_{b,j}$.
5. Form $\Delta\mathbf{Z}_b$ matrix by pairwise differencing \mathbf{Z}_b .
6. Form $\Phi_b := ((\Delta\mathbf{Z}_b)'(\Delta\mathbf{Z}_b))^{-1}$.
7. Form $\mathbf{Q}_b := (\sqrt{\Omega_b}) \circ (\Delta\mathbf{Z}_b)$.
8. Solve the following constrained optimization problem. Note that the first set of constraints, $\mathbf{0} = \boldsymbol{\pi}(\boldsymbol{\xi}; \boldsymbol{\Theta}) - \mathbf{s}$, are intact from the original problem.

$$\begin{aligned}
 & \min_{(\boldsymbol{\Theta}, \boldsymbol{\xi}, \mathbf{g})} \mathbf{g}' \Phi_b \mathbf{g} \\
 \text{s.t. } & \mathbf{0} = \boldsymbol{\pi}(\boldsymbol{\xi}; \boldsymbol{\Theta}) - \mathbf{s} \\
 & \mathbf{0} = \mathbf{Q}_b'(\Delta\boldsymbol{\xi}_b) - \mathbf{g},
 \end{aligned} \tag{A.2}$$

The constraint Jacobian is:

$$\begin{pmatrix} \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\Theta}} & \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{Q}_b'(\Delta\boldsymbol{\xi}_b)}{\partial \boldsymbol{\xi}} & -\mathbf{I} \end{pmatrix}.$$

The main change of (A.2) from point estimation problem (A.1) is the second part of the constraint and its Jacobian component. Recall $\mathbf{Q}_b := (\sqrt{\boldsymbol{\Omega}_b}) \circ (\Delta \mathbf{Z}_b)$ and the structure of $\Delta \boldsymbol{\xi}_b$:

$$\Delta \boldsymbol{\xi}_b := \begin{pmatrix} \xi_{b_1} - \xi_{b_2} \\ \xi_{b_1} - \xi_{b_3} \\ \vdots \\ \xi_{b_1} - \xi_{b_n} \\ \xi_{b_2} - \xi_{b_3} \\ \xi_{b_2} - \xi_{b_4} \\ \vdots \\ \xi_{b_2} - \xi_{b_n} \\ \vdots \\ \xi_{b_{n-1}} - \xi_{b_n} \end{pmatrix}.$$

The difference from point estimation is that the pairwise differencing is taken in the order with respect to the bootstrap shuffle indicator b_1, b_2, \dots, b_n . However, $\boldsymbol{\Psi}_b$ should be filled with the derivatives of $\mathbf{Q}'_b(\Delta \boldsymbol{\xi}_b)$ with respect to the original order of $(\xi_1, \xi_2, \dots, \xi_n)$. Furthermore, there can be some rows of $\boldsymbol{\Psi}_b$ such that it is a zero vector due to the resampling scheme.

The reshuffling scheme described above makes the calculation of $\boldsymbol{\Psi}_b$ slightly more complicated than the calculation of $\boldsymbol{\Psi}$. Consider $\boldsymbol{\psi}'_{b,j}$, j 'th row of $\boldsymbol{\Psi}_b = \left(\frac{\partial \mathbf{Q}'_b(\Delta \boldsymbol{\xi}_b)}{\partial \boldsymbol{\xi}} \right)'$. If $\boldsymbol{\xi}_b$ does not contain ξ_j , $\boldsymbol{\psi}'_{b,j} = \mathbf{0}$. Now suppose $\boldsymbol{\xi}_b$ contains only one ξ_j . Then, there is going to be one b_k such that b_k th element of \mathbf{n}_b is j . Now, for such j, k pair, $\boldsymbol{\psi}'_{b,j}$ becomes

$$\boldsymbol{\psi}'_{b,j} = \mathbf{z}'_{b_k} \left(\sum_{l=1}^n \sqrt{\hat{\omega}_{b_k, b_l}} \right) - \sum_{l=1}^n \left(\sqrt{\hat{\omega}_{b_k, b_l}} \mathbf{z}'_{b_l} \right),$$

where \mathbf{z}'_{b_k} is k th row of \mathbf{Z}_b . If there are more than one b_k such that b_k th element of \mathbf{n}_b is j ,

$$\boldsymbol{\psi}'_{b,j} = \sum_{\{b_k: b_k \text{th element of } \mathbf{n}_b \text{ is } j\}} \left\{ \mathbf{z}'_{b_k} \left(\sum_{l=1}^n \sqrt{\hat{\omega}_{b_k, b_l}} \right) - \sum_{l=1}^n \left(\sqrt{\hat{\omega}_{b_k, b_l}} \mathbf{z}'_{b_l} \right) \right\}.$$

A.4 Bootstrap Algorithm for Random-Coefficients Logit Demand Estimation When All the Observed Shares are Positive

When all the observed shares are positive, we have the usual MPEC formulation of the problem:

$$\min_{(\Theta, \xi, \mathbf{g})} \mathbf{g}' \Phi \mathbf{g}$$

$$s.t. \quad \mathbf{0} = \pi(\xi; \Theta) - \mathbf{s} \tag{A.3}$$

$$\mathbf{0} = \mathbf{Z}' \xi - \mathbf{g}. \tag{A.4}$$

For each bootstrap trial, Φ should be replaced with Φ_b and $\mathbf{0} = \mathbf{Z}' \xi - \mathbf{g}$ should be replaced with $\mathbf{0} = \mathbf{Z}'_b \xi_b - \mathbf{g}$, where $(\Phi_b, \mathbf{Z}_b, \xi_b)$ denotes the bootstrap-reshuffled matrices/vectors of (Φ, \mathbf{Z}, ξ) . However, the market-share equation (A.3) should be left intact.

The gradient of the objective function is:

$$\nabla \mathbf{g}' \Phi_b \mathbf{g} = \begin{pmatrix} \mathbf{0}' & \mathbf{0}' & 2\mathbf{g}' \Phi'_b \end{pmatrix}.$$

The constraint Jacobian of the problem is

$$\begin{pmatrix} \frac{\partial \pi}{\partial \Theta} & \frac{\partial \pi}{\partial \xi} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}'_b & -\mathbf{I} \end{pmatrix},$$

and the Hessian of the Lagrangian is identical with the the original problem, except for the second derivative of the objective function which is $2\Phi_b$.

A.5 Choice of Kernels and Bandwidth

For the kernel $H(\cdot)$ in (4.4), we use the product kernel with the marginal kernel for the continuous elements of \mathbf{w}_j the Gaussian kernel. For the discrete elements of \mathbf{w}_j , we use the kernel described in pages 136-137 of Li and Racine (2008). Silverman's rule is used for the bandwidth of Gaussian marginal kernel, and vector of ones are used for the bandwidth of discrete marginal kernel. The final results are robust to the choice of the bandwidth of Stage 2. For the kernel $\kappa(\cdot)$ and bandwidth h_{n_d} in (4.7), we use the Gaussian kernel again with Silverman's rule-of-thumb bandwidth.

B Sensitivity of Estimates on the Imputed Magnitude of Zero Shares

In this section, we present the sensitivity analysis of the model parameter estimates with respect to the magnitude of the small number being imputed in place of zero shares, estimation results both for our Monte Carlo datasets and our real-world dataset. We estimated the Homogeneous Logit model of demand with instrumenting for the prices, and varying the magnitudes of the number being imputed from 10^{-5} to 10^{-300} .

The datasets being used here are identical to the datasets used in sections 5 and 6, respectively. Notice that the estimates of Impute 10^{-12} in Table 9 are identical to those of Impute 10^{-12} of Table 7. Analogously, the estimates of Impute 10^{-12} in Table 8 are identical to those of Impute 10^{-12} in Table 5.

Table 7: Sensitivity of the Estimates to the Magnitude of the Small Number Being Imputed: Monte Carlo Datasets

(a) With Selection-on-Unobservables														
DGP	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute		
	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-12}	10^{-15}	10^{-20}	10^{-30}	10^{-50}	10^{-100}	10^{-200}	10^{-300}
$\bar{\alpha}$ [-2]	-3.605	-4.474	-5.344	-6.213	-7.083	-7.953	-9.692	-12.301	-16.649	-25.346	-42.738	-86.220	-173.184	-260.148
$\bar{\beta}_1$ [1]	0.591	0.591	0.591	0.591	0.591	0.591	0.590	0.590	0.590	0.589	0.587	0.582	0.573	0.563
$\bar{\beta}_2$ [2]	1.225	1.222	1.220	1.217	1.214	1.211	1.205	1.197	1.183	1.154	1.097	0.955	0.670	0.385
$\bar{\beta}_3$ [2]	1.236	1.238	1.239	1.241	1.242	1.244	1.247	1.251	1.259	1.274	1.303	1.378	1.527	1.676

(b) Without Selection-on-Unobservables														
DGP	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute	Impute
	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-12}	10^{-15}	10^{-20}	10^{-30}	10^{-50}	10^{-100}	10^{-200}	10^{-300}
$\bar{\alpha}$ [-2]	-0.934	-0.924	-0.915	-0.905	-0.895	-0.886	-0.866	-0.837	-0.789	-0.692	-0.498	-0.013	0.956	1.926
$\bar{\beta}_1$ [1]	0.562	0.556	0.550	0.543	0.537	0.530	0.518	0.498	0.467	0.403	0.275	-0.044	-0.682	-1.320
β_2 [2]	1.160	1.150	1.141	1.132	1.122	1.113	1.094	1.067	1.020	0.927	0.741	0.275	-0.656	-1.588
$\bar{\beta}_3$ [2]	1.163	1.154	1.146	1.137	1.128	1.119	1.101	1.075	1.031	0.942	0.766	0.324	-0.560	-1.444

Notes. Column names indicate the imputed share value used to handle zeros during the estimation. True parameter values are reported in square brackets. The reported estimates are the means of each parameter across the 100 replications of the data generating process. Standard deviations across converged trials are reported in the parentheses.

Table 8: Sensitivity of the Estimates to the Magnitude of the Small Number Being Imputed: Real-world Ice Cream Data

	10 ⁻⁴	10 ⁻⁵	10 ⁻⁶	10 ⁻⁷	10 ⁻⁸	10 ⁻⁹	10 ⁻¹⁰	10 ⁻¹²	10 ⁻¹⁵	10 ⁻²⁰	10 ⁻³⁰	10 ⁻⁵⁰	10 ⁻¹⁰⁰	10 ⁻²⁰⁰	10 ⁻³⁰⁰
Mean Elast.	-3.392	-5.171	-6.950	-8.729	-10.509	-12.288	-14.067	-17.625	-22.963	-31.858	-49.650	-85.233	-174.19	-352.105	-530.02
Price	-0.917 (0.134)	-1.399 (0.205)	-1.880 (0.276)	-2.361 (0.347)	-2.842 (0.418)	-3.324 (0.489)	-3.805 (0.560)	-4.767 (0.702)	-6.211 (0.915)	-8.617 (1.270)	-13.430 (1.981)	-23.054 (3.402)	-47.116 (6.954)	-95.239 (14.060)	-143.362 (21.165)
1000 Cal/pack	-0.145 (0.432)	-0.246 (0.666)	-0.347 (0.900)	-0.448 (1.136)	-0.549 (1.371)	-0.650 (1.606)	-0.751 (1.842)	-0.953 (2.312)	-1.256 (3.019)	-1.762 (4.196)	-2.773 (6.552)	-4.794 (11.263)	-9.848 (23.040)	-19.956 (46.595)	-30.064 (70.151)
Vanilla	0.536 (0.048)	0.801 (0.073)	1.065 (0.098)	1.329 (0.123)	1.594 (0.148)	1.858 (0.173)	2.122 (0.198)	2.651 (0.248)	3.444 (0.324)	4.766 (0.450)	7.409 (0.701)	12.695 (1.205)	25.912 (2.463)	52.345 (4.980)	78.778 (7.496)
Chocolate	0.153 (0.044)	0.227 (0.067)	0.300 (0.091)	0.374 (0.115)	0.447 (0.139)	0.521 (0.162)	0.594 (0.186)	0.741 (0.234)	0.962 (0.305)	1.329 (0.424)	2.064 (0.663)	3.535 (1.139)	7.211 (2.33)	14.562 (4.713)	21.914 (7.095)
Yogurt	-0.470 (0.045)	-0.721 (0.069)	-0.972 (0.094)	-1.222 (0.119)	-1.473 (0.144)	-1.724 (0.169)	-1.974 (0.193)	-2.476 (0.243)	-3.227 (0.317)	-4.481 (0.442)	-6.987 (0.69)	-12.000 (1.186)	-24.532 (2.427)	-49.597 (4.908)	-74.662 (7.390)
Mint	0.146 (0.067)	0.229 (0.103)	0.311 (0.139)	0.393 (0.175)	0.476 (0.211)	0.558 (0.248)	0.641 (0.284)	0.806 (0.356)	1.053 (0.465)	1.465 (0.647)	2.289 (1.010)	3.938 (1.735)	8.059 (3.550)	16.301 (7.179)	24.543 (10.808)
Fruit	-0.034 (0.059)	-0.045 (0.092)	-0.056 (0.124)	-0.068 (0.157)	-0.079 (0.190)	-0.090 (0.222)	-0.101 (0.255)	-0.124 (0.320)	-0.157 (0.418)	-0.214 (0.582)	-0.326 (0.908)	-0.550 (1.562)	-1.112 (3.195)	-2.234 (6.462)	-3.357 (9.729)
Breyers	-0.318 (0.152)	-0.493 (0.232)	-0.669 (0.313)	-0.844 (0.393)	-1.019 (0.474)	-1.194 (0.555)	-1.370 (0.635)	-1.720 (0.796)	-2.246 (1.038)	-3.122 (1.442)	-4.875 (2.249)	-8.381 (3.862)	-17.144 (7.897)	-34.671 (15.966)	-52.198 (24.034)
Deans	-0.912 (0.192)	-1.395 (0.293)	-1.877 (0.394)	-2.360 (0.495)	-2.842 (0.597)	-3.325 (0.698)	-3.807 (0.799)	-4.772 (1.002)	-6.220 (1.307)	-8.632 (1.815)	-13.458 (2.830)	-23.108 (4.861)	-47.233 (9.939)	-95.483 (20.094)	-143.733 (30.250)
Edys	-0.770 (0.168)	-1.172 (0.257)	-1.575 (0.347)	-1.978 (0.436)	-2.381 (0.526)	-2.783 (0.615)	-3.186 (0.705)	-3.991 (0.884)	-5.199 (1.153)	-7.213 (1.601)	-11.240 (2.498)	-19.293 (4.291)	-39.428 (8.774)	-79.697 (17.74)	-119.966 (26.707)

Notes. Column names indicate the imputed share value used to handle zeros during the estimation. Bootstrap standard errors for each coefficient estimate are reported in parentheses. *Mean Elast.* row indicates the mean own-price elasticities across all the observations.

C Homogeneous Logit Demand Model Estimates on Monte Carlo Datasets

Table 9 presents the estimation results for the Homogeneous Logit models using the same 100 Monte Carlo datasets generated and used in section 5. It turns out that not accounting for the random-coefficients during the estimation biases the coefficient estimates slightly. Even without random-coefficients, our PDWGMM estimator performs very well compared to the conventional methods to handle zeros when selection-on-unobservables affect the prevalence of zero market shares.

Table 9: Homogeneous Logit Monte Carlo Data Estimation Results

DGP	(a) With Selection-on-Unobservables				(b) Without Selection-on-Unobservables				(c) No Zero Shares	
	PDWGMM	Heckman	Drop 0	Impute 10^{-12}	PDWGMM	Heckman	Drop 0	Impute 10^{-12}	PDWGMM	Drop 0
Homogeneous Logit										
$\bar{\alpha}$ [-2]	-2.225	-1.797	-4.429	-9.692	-1.756	-1.679	-1.668	-0.866	-1.752	-1.673
$\bar{\beta}_1$ [1]	1.146	1.030	0.999	0.590	1.047	1.005	1.002	0.518	1.036	0.997
$\bar{\beta}_2$ [2]	2.417	2.139	2.077	1.205	2.139	2.061	2.051	1.094	2.145	2.066
$\bar{\beta}_3$ [2]	2.415	2.139	2.085	1.247	2.148	2.065	2.055	1.101	2.143	2.064

Notes. Column names indicate how the zero shares are handled during the estimation. True parameter values are reported in square brackets. The reported estimates are the means of each parameter across the 100 replications of the data generating process. *Heckman* is the classical Heckman selection estimator with inverse Mills ratio as an additional regressor.