Quantitative Economics 14 (2023), 381–418 1759-7331/20230381

Estimating demand for differentiated products with zeroes in market share data

Amit Gandhi
Airbnb

Zhentong Lu
Bank of Canada

Xiaoxia Shi
UW-Madison

In this paper, we introduce a new approach to estimating differentiated product demand systems that allows for products with zero sales in the data. Zeroes in demand are a common problem in differentiated product markets, but fall outside the scope of existing demand estimation techniques. We show that with a lower bound imposed on the expected sales quantities, we can construct upper and lower bounds for the conditional expectation of the inverse demand. These bounds can be translated into moment inequalities that are shown to yield consistent and asymptotically normal point estimators for demand parameters under natural conditions. In Monte Carlo simulations, we demonstrate that the new approach works well even when the fraction of zeroes is as high as 95%. We apply our estimator to supermarket scanner data and find that correcting the bias caused by zeroes has important empirical implications, for example, price elasticities become twice as large when zeroes are properly controlled.

Keywords. Demand estimation, differentiated products, measurement error, moment inequality, zero.

JEL classification. C01, C12, L10, L81.

Amit Gandhi: amit.gandhi@airbnb.com
Zhentong Lu: zlu@bankofcanada.ca
Xiaoxia Shi: xshi@ssc.wisc.edu

A previous version of this paper was circulated under the title "Estimating demand for differentiated products with error in market shares." We are thankful to Steven Berry, Jean-Pierre Dubé, Philip Haile, Bruce Hansen, Ulrich Müller, Aviv Nevo, Jack Porter, Dan Quint, and Chris Taber for insightful discussions and suggestions; We would also like to thank the participants at the 2011 MIT Econometrics of Demand Conference, Chicago-Booth Marketing Lunch, the 2012 Northwestern Conference on "Junior Festival on New Developments in Microeconometrics," the 2012 Cowles Foundation Conference on "Structural Empirical Microeconomic Models," 3rd Cornell—Penn State Econometrics and Industrial Organization Workshop (2016), the IAAE Annual Conference in 2021, as well as seminar participants at Wisconsin-Madison, Wisconsin-Milwaukee, Cornell, Indiana, Princeton, NYU, Penn, the Federal Trade Commission, and UCSD for their many helpful comments and questions. Xiaoxia Shi dedicates this paper to the loving memory of her mother, Guizhi Ma, who passed away in a car crash in December 2019.

 https://doi.org/10.3982/QE1593

1. Introduction

In this paper, we introduce a new approach to differentiated product demand estimation that allows for zeroes in empirical market share data. Such zeroes are a highly prevalent feature of demand in a variety of empirical settings, ranging from workhorse retail scanner data, to data as diverse as homicide rates and international trade flows (we discuss these examples in further depth below). Zeroes naturally arise in "big data" applications, which allow for increasingly granular views of consumers, products, and markets (see, e.g., Quan and Williams (2018), and Nurski and Verboven (2016)). Unfortunately, the standard estimation procedures using inverse demand function following the seminal Berry, Levinsohn, and Pakes (1995) (BLP for short) cannot be used in the presence of zero empirical shares—the inverse demand is simply not well-defined at zeroes. Furthermore, ad hoc fixes to market zeroes that are sometimes used in practice, such as dropping zeroes from the data or replacing them with small positive numbers, are subject to biases, which can be quite large (because the slope of the inverse demand is arbitrarily large around zero). This has left empirical work on demand for differentiated products without satisfying solutions to the zero shares problem, and often force researchers to aggregate their rich data on naturally defined products to crude artificial products, which limits the type of questions that can be answered. This is the key problem that our paper aims to solve.

In this paper, we provide an approach to estimating differentiated product demand models that provides consistency and asymptotic normality for demand parameters despite a possibly large presence of zero market shares in the data. We start by noting that the zeroes are caused by the wedge between the empirical shares (s_{jt}) and the true choice probabilities (π_{jt}): while the latter is always positive, the former can be zero because of sample noise. We show how the zeroes in empirical shares may not simply be a data anomaly, but an essential feature of markets with a rich product variety, even if the number of consumers (n_t) is large. By market design, expected sales ($n_t\pi_{jt}$) of some products do not increase with n_t, and as a result, their empirical market shares are zero with nonvanishing probabilities. We then show that by imposing a lower bound to the expected sales we can construct upper and lower bounds for the conditional expectation of the inverse demand. The bounds are used to construct a set of moment *inequalities*, which are valid in the presence of the zeroes, and more generally in the presence of sampling error in market shares.[1]

The moment inequalities can be directly used for parameter inference with the help of set inference methods in the econometrics literature but for computational reasons,

[1] In the last couple of years, new aggregate demand models have been considered that accommodate zeroes in market share data in Dube, Hortacsu, and Joo (2020) and Lima (2021). Dube et al. model the products with zero market shares as ones that are not in any consumer's consideration set. Lima's model rationalizes the zeros in market shares by restricting the support of the idiosyncratic taste shock. Neither paper deals with the sample noise issue in observed market shares. Since Dube et al., Lima, and our paper's methods rely on nonnested assumptions on the source of zeros, in practice, knowing the true source of zero is important for choosing the appropriate method. When in doubt, it is advisable to implement multiple methods and compare the results. A potentially interesting direction for future research is to combine those methods into a more generally applicable solution to the problem of zero market shares.

we give a point-identification condition and propose a point estimator instead. We show that our point estimator is consistent so long as n_t is large and there is an exogenous product or market characteristic, or a group of them, that can identify a positive mass of observations whose latent choice probabilities are bounded sufficiently away from zero, for example, product-market pairs for whom the observed market shares are not likely to be zero. This is natural in many applications (as illustrated in Section 5), and strictly generalizes the restrictions on choice probabilities for consistency under the traditional approach. Asymptotic normality then follows by similar arguments as those for censored regression models by Kahn and Tamer (2009).

Computationally, our estimator closely resembles the traditional approach with only a slight adjustment in how the empirical moments are constructed. In particular, it is no more burdensome than the usual estimation procedures for BLP and can be implemented using either the standard nested fixed-point method of the original BLP, or the MPEC method as advocated more recently by Dubé, Fox, and Su (2012).

We investigate the finite sample performance of the approach in a variety of mixed logit examples. We find that our estimator works well even when the the fraction of zeros is as high as 95%, while the standard procedure with the observations with zeroes deleted yields severely biased estimators even with mild or moderate fractions of zeroes.

We apply our bounds approach to widely used scanner data from the Dominicks Finer Foods (DFF) retail chain. In particular, we estimate demand for the tuna category as previously studied by Chevalier, Kashyap, and Rossi (2003) and continued by Nevo and Hatzitaskos (2006) in the context of testing the loss-leader hypothesis of retail sales. We find that controlling for products with zero demand using our approach gives demand estimates that can be more than twice as elastic than standard estimates that select out the zeroes. We also show that the estimated price elasticities increase substantially during Lent (a high demand period for this product category) after we control for the zeroes. Both of these findings have implications for reconciling the loss-leader hypothesis with the data.

The plan of the paper is the following. In Section 2, we illustrate the stylized empirical pattern of Zipf's law where market zeroes naturally arise. In Section 3, we describe our solution to the zeroes problem using a simple logit setup without random coefficients to make the essential matters transparent. In Section 4, we extend the moment inequality construction and our estimator to general discrete choice model possibly with random coefficients. Section 5 discusses the point-identification condition. Sections 6 and 7 present the theoretical properties of the proposed estimator. Section 8 presents results of Monte Carlo simulations and Section 9 presents the application to the DFF data, respectively. Section 10 concludes.

2. Market zeroes

In this section, we highlight the empirical pattern of zeroes. Here, we primarily use workhorse store level scanner data to illustrate these patterns. It is the same data that will also be used for our application. However, we emphasize that our focus here on scanner data is only for the sake of a concrete illustration of the market zeroes problem—the key patterns we highlight in scanner data are also present in many other economic settings where demand estimation techniques are used (discussed further below and illustrated in Section A of the Supplemental Appendix (Gandhi, Lu, and Shi (2023)).

We employ here a widely-studied store level scanner data set from the Dominick's Finer Foods grocery chain, which is a public data set that has been used by many researchers.[2] The data comprise 93 Dominick's Finer Foods stores in the Chicago metropolitan area over the years from 1989 to 1997. Like other store level scanner data sets, this data set provides demand information (price, sales, marketing) at the store/week/UPC level, where a universal product code (UPC) is a unique bar code that identifies a natural product.[3]

Table 1 presents information on the resulting product variety across the different product categories in the data. The first column shows the number of products in an average store/week—the number of UPCs can be seen varying from roughly fifty (e.g., bath tissue) to over four hundred (e.g., soft drinks) within even these narrowly defined categories. Thus, there is considerable product variety in the data. The next two columns illustrate an important aspect of this large product variety: there are often just a few UPCs that dominate each product category whereas most UPCs are not frequently chosen. The second column illustrates this pattern by showing the well-known "80/20" rule that prevails in our data: we see that roughly 80% of the total quantity purchased in each category is driven by the top 20% of the UPCs in the category. In contrast to these "top sellers," the other 80% of UPCs contain relatively "sparse sellers" that share the remaining 20% of the total volume in the category. The third column shows an important consequence of this sparsity: many UPCs in a given week at a store simply do not sell. In particular, we see that the fraction of observations with zero sales can even be nearly 60% for some categories.

We can visualize this situation in another way by fixing a product category (here we use canned tuna) and simply plotting the histogram of the volume sold for each week/UPC realization for a single store in the data. This frequency plot is given in Figure 1. As can be seen, there is a sharp decay in the empirical frequency as the purchase quantity becomes larger, with a long thin tail.[4] In particular, the bulk of UPCs in the store has small purchase volume: the median UPC sells less than 10 units a week, which is less than 1.5% of the median volume of tuna the store sells in a week. The mode of the frequency plot is a zero share.

This power-law decay in the frequency of product demand is often associated with "Zipf's law" or the "the long tail," which has a long history in empirical economics.[5]

[2]For a complete list of papers using this data set, see the website of Dominick's Database: https://www.chicagobooth.edu/research/kilts/datasets/dominicks

[3]Store level scanner data can often be augmented with a panel of household level purchases (available, e.g., through IRI or Nielsen). Although the DFF data do not contain this micro level data, the main points of our analysis are equally applicable to the case where household level data is available. Store level purchase data can be viewed as a special case household level data where all households are observationally identical (no observable individual level characteristics).

[4]We plot the long tail pattern differently from a commonly seen illustration of power law using rank-size distribution ("size against rank or popularity"), but the difference is only cosmetic (basically flipping the x- and y-axis); the two ways of plotting convey the same information.

[5]See Anderson (2006) for a historical summary of Zipf's law and many examples from the social and natural sciences. See Gabaix (1999) for an application of Zipf's law to the economics literature.

TABLE 1. Selected product categories in the Dominick's Database.

Category	Average Number of UPCs in a Store/Week Pair	Percent of Total Sale of the Top 20% UPCs	Percent of Zero Sales
Beer	179	87.18%	50.45%
Cereals	212	72.08%	27.14%
Crackers	112	81.63%	37.33%
Dish detergent	115	69.04%	42.39%
Frozen dinners	123	66.53%	38.32%
Frozen juices	94	75.16%	23.54%
Laundry detergents	200	65.52%	50.46%
Paper towels	56	83.56%	48.27%
Refrigerated juices	91	83.18%	27.83%
Soft drinks	537	91.21%	38.54%
Snack crackers	166	76.39%	34.53%
Soaps	140	77.26%	44.39%
Toothbrushes	137	73.69%	58.63%
Canned tuna	118	82.74%	35.34%
Bathroom tissues	50	84.06%	28.14%

We present further illustrations of this long-tail demand pattern found in international trade flows as well as cross-county homicide rates in Supplemental Appendix A, which provide a sense of the generality of these stylized facts.

The key takeaway from these illustrations is that the presence of market zeroes in the data is closely intertwined to the prevalence of power-law patterns of demand. We will exploit this relationship to place structure on the data generating process that underlies market zeroes.

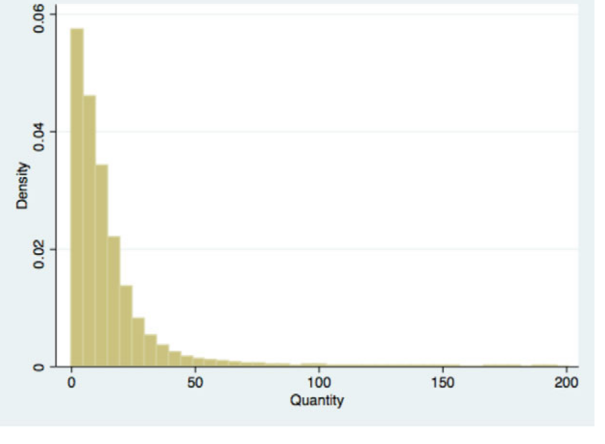

FIGURE 1. Zipf's law in scanner data.

3. A first pass through logit demand

Why do zero shares create a problem for demand estimation? In this section, we use the workhorse multinomial logit model to explain the zeroes problem and our solution. The general case is treated in the next section. In both cases, we assume that the econometrician observes a data set of $\{(n_t, s_{jt}, x_{jt}) : j = 1, \dots, J_t, t = 1, \dots, T\}$, where n_t is the number of potential consumers in market t, s_{jt} is the fraction of those consumers choosing product j, and x_{jt} is the vector of observed characteristics of the product j and/or market t that often includes price, J_t is the number of inside products in market t, and T is the number of markets. We focus on the case where there are many markets.

3.1 *Making sense of the zeroes*

Consider a multinomial logit model for the demand of J_t products ($j = 1, \dots, J_t$) and an outside option ($j = 0$). A consumer i derives utility $u_{ijt} = \delta_{jt} + \epsilon_{ijt}$ from product j in market t, where δ_{jt} is the mean utility of product j in market t, and ϵ_{ijt} is the idiosyncratic taste shock that follows the type-I extreme value distribution. As is standard, the mean utility δ_{jt} of product $j > 0$ is modeled as

$$\delta_{jt} = x'_{jt}\beta_0 + \xi_{jt}, \tag{1}$$

where ξ_{jt} is the unobserved characteristic. The outside good $j = 0$ has mean utility normalized to $\delta_{0t} = 0$. The parameter of interest is β_0.

Each consumer chooses the product that yields the highest utility:

$$s_{ijt} = 1\{u_{jt} \geq u_{j't}\ \forall j' = 0, 1, \dots, J_t\}, \quad \text{for } j = 0, 1, \dots, J_t. \tag{2}$$

Aggregating consumers' choices, we obtain the true choice probability of product j in market t, denoted as

$$\pi_{jt} = \Pr(\text{product } j \text{ is chosen in market } t) = E[s_{ijt}|\delta_{1t}, \dots, \delta_{J_t t}].$$

The standard approach introduced by Berry (1994) for estimating β_0 is to combine demand system inversion and instrumental variables.

First, for demand inversion, one uses the logit structure to find that

$$\delta_{jt} = \log(\pi_{jt}) - \log(\pi_{0t}), \quad \text{for } j = 1, \dots, J_t. \tag{3}$$

To handle the potential endogeneity of x_{jt} (i.e., its correlation with ξ_{jt}), one finds some excluded instruments, which along with the exogenous controls in x_{jt} form z_{jt} such that

$$E[\xi_{jt}|z_{jt}] = 0. \tag{4}$$

Then two stage least squares with δ_{jt} defined in (3) as the dependent variable becomes the identification strategy for β_0.

Unfortunately, π_{jt} is not observed as data—it is a theoretical choice probability defined by the model but only indirectly revealed through actual consumer choices. The standard approach to this following Berry (1994), Berry, Levinsohn, and Pakes (1995), and many subsequent papers in the literature has been to substitute s_{jt} the empirical market share for π_{jt}, where

$$s_{jt} = n_t^{-1} \sum_{i=1}^{n_t} s_{ijt} \quad \text{for } j = 0, 1, \dots, J_t, \tag{5}$$

and run a two-stage least square with $\log(s_{jt}) - \log(s_{0t})$ as dependent variable, x_{jt} as covariates, and z_{jt} as instruments to obtain estimates for β_0. The theoretical justification used in the literature assume that n_t is large, and importantly, π_{jt} either is bounded away from zero or converges to zero at a slower rate than $1/n_t$. Under these assumptions, Berry, Linton, and Pakes (2004) and Freyberger (2015) show that plugging in s_{jt} for π_{jt} at worst leads to a correctible bias.

However, for data sets with the power law pattern described in Section 2, a large proportion of the s_{jt}'s are zeroes. Substituting s_{jt} for π_{jt} is no longer feasible, and the theoretical assumptions used to justify that practice are no longer compatible with the data. The former is because $\log(0)$ is not finite, and the latter is because under the assumption that π_{jt} approaches zero at a slower rater than $1/n_t$. We have $\Pr(s_{jt} = 0) \to 0$, which is not consistent with the large number of zeroes in the data.

We rationalize the large number of zeros in s_{jt} at seemingly large n_t by allowing π_{jt} to approach zero at the rate of $1/n_t$. When π_{jt} approaches zero at this rate, for example, $\pi_{jt} = c/n_t$ for a constant $c > 0$, we have

$$\lim_{n_t \to \infty} \Pr(s_{jt} = 0) = \lim_{n_t \to \infty} (1 - c/n_t)^{n_t} = \exp(-c). \tag{6}$$

Thus, zeroes arise naturally in this framework. In our bound construction below, we will assume a much weaker lower bound on π_{jt} than the existing literature: $\pi_{jt} \geq \underline{\varepsilon}_1/n_t$ for some fixed constant $\underline{\varepsilon}_1$.

There is a simple supply side explanation for why the choice probability of some products should approach zero at the exact rate of $1/n_t$ and why there may be a lower bound for $n_t\pi_{jt}$. A market with the power-law feature described in Section 2 may be thought of as one with a few dominant products that coexist with a competitive fringe (see, e.g., Shimomura and Thisse (2012)). The fringe products enjoy free entry and exit and are subject to a fixed cost, denoted f_{jt}. The free entry and exit drives their expected profit to zero:

$$n_t \pi_{jt} m_{jt} - f_{jt} = 0, \tag{7}$$

where m_{jt} is the average mark-up. Then $n_t\pi_{jt} = f_{jt}/m_{jt}$. And $\pi_{jt} \geq \underline{\varepsilon}_1/n_t$ holds for some $\underline{\varepsilon}_1$ if there are a lower bound for f_{jt} and an upper bound for m_{jt}.[6] If there is also an upper bound for f_{jt} and a lower bound for m_{jt}, then π_{jt} approaches zero at the rate of $1/n_t$. The existence of such bounds is reasonable in differentiated product markets.[7]

[6]The calculation assumes single-product firms. Multiproduct firms stop putting out new products sooner because they internalize the business stealing effect of new products on their existing products.

[7]The only bound that might be disputable is the lower bound for the average markup because markup is endogenous. But even that has some supporting evidence in the literature: Armstrong (2016) shows that the markup converges to a positive constant rather than zero when the number of firms grows to infinity.

3.2 *Estimation problem with zeroes*

As mentioned above, the zeroes pose an immediate challenge to estimation: $\log(s_{jt})$ is $-\infty$ when $s_{jt}=0$. This makes the standard BLP estimator ill-defined. A common workaround is to ignore the (jt)'s with $s_{jt}=0$, effectively lumping those j's into the outside option in market t. This however leads to a selection problem. To see this, suppose $s_{jt}=0$ for some (jt) and one drops these observations from the analysis—effectively one is using a selected sample where the selection criterion is $s_{jt}>0$. In this selected sample, the conditional mean of ξ_{jt} is no longer a constant. This is the well-known selection-on-unobservables problem and with such sample selection, an attenuation bias ensues.[8] The attenuation bias generally leads to demand estimates that appear to be too inelastic.[9]

Another commonly adopted empirical "trick" is to add a small positive number $\epsilon>0$ to the s_{jt}'s that are zero, and use the resulting modified shares $s^{\epsilon}_{jt}>0$ in place of π_{jt}.[10] However, this trick only treats the symptom, that is, $s_{jt}=0$, but overlooks the nature of the problem: the true choice probability π_{jt} is small. And in this case, small estimation error in any estimator $\widehat{\pi}_{jt}$ of π_{jt} would lead to large error in the plugged-in version of δ_{jt} and the estimation of β_0. This problem manifests itself directly because the estimate $\widehat{\beta}$ can be incredibly sensitive to the particular choice of the small number being added and there is little guidance on what is the "right" choice of the small number. In general, like selecting away the zeroes, the "adding a small number trick" is also a biased estimator for β_0. We illustrate both biases in the Monte Carlo section (Section 8).

Despite their failure as general solutions, these "ad hoc zero fixes" have in them what could be a useful idea. Perhaps the variation among the nonzero share observations can be used to estimate the model parameters, while at the same time the presence of zeroes is controlled in such a way that avoids bias. We will present a new estimator that formalizes this possibility by using moment *inequalities* to control for the zeroes in the data while using the variation in the remaining part of the data to estimate the demand parameters.

---

[8] To see why $E[\xi_{jt}|x_{jt},s_{jt}>0]$ is not a constant, consider two values of x_{jt}: x, x^* such that $x'\beta>x^{*\prime}\beta$, and consider the homoskedastic case for simplicity. For each given value of x_{jt}, the criterion $s_{jt}>0$ selects high values of ξ_{jt} and leaves out low values of ξ_{jt}. Moreover, the selection is more severe for x^* than for x because the unobservable (to econometricians) needs to more appealing to induce a positive observed market share when the observable characteristic is less appealing.

Thus, we should have

$$E\big[\xi_{jt}|x_{jt}=x^*,s_{jt}>0\big]>E[\xi_{jt}|x_{jt}=x,s_{jt}>0], \tag{8}$$

and clearly, $E[\xi_{jt}|x_{jt},s_{jt}>0]$ is not a constant.

[9] It is easy to see that the selection bias is of the same direction if the selection criterion is instead $s_{jt}>0$ for all t, as one is effectively doing when focusing on a few top sellers that never demonstrate zero sales in the data. The reason is that the event $s_{jt}>0$ for all t contains the event $s_{jt}>0$ for a particular t. If the markets (ξ_{jt}'s) are independent, the particular t part of the selection dominates.

[10] Berry, Linton, and Pakes (2004) and Freyberger (2015) study the biasing effect of plugging in s_{jt} for π_{jt}. Their bias corrections do not apply when there are zeroes in the empirical shares.

3.3 *Constructing moment inequalities*

Our approach builds on two estimators of $\log(\pi_{jt})$. We refer to them as the upper and lower bounds of $\log(\pi_{jt})$ because they bound $\log(\pi_{jt})$ from above and below *on average* in the sense discussed below. These bounds are

$$\log\bigl((n_t s_{jt} + \iota_u)/n_t\bigr) \quad \text{and} \quad \log\bigl((n_t s_{jt} + \iota_\ell)/n_t\bigr), \tag{9}$$

where ι_u and ι_ℓ are two positive numbers that we now construct.

To construct ι_u and ι_ℓ, note that $n_t s_{jt}$ follows a binomial distribution given n_t and π_{jt}: $Bin(n_t, \pi_{jt})$.[11] For each fixed n and π, and $\iota \geq 0$, define the function

$$f(\iota; n, n\pi) := E\bigl[\log(n_t s_{jt} + \iota) - \log(n_t \pi_{jt}) | n_t = n, \pi_{jt} = \pi\bigr].$$

The function f is negative infinity at $\iota = 0$ (because s_{jt} can be 0 with positive probability), strictly increasing with ι, and approaches positive infinity as $\iota \to \infty$. Therefore, at each n and π, the function crosses zero once and only once. We let the point of crossing be denoted $\iota^*(n, n\pi)$, which is defined implicitly by the equation:

$$f\bigl(\iota^*(n, n\pi); n, n\pi\bigr) = 0. \tag{10}$$

This quantity can be calculated because the function $f(\iota; n, n\pi)$ (i.e., the expectation) can be calculated using the binomial distribution.

As explained in Section 3.1 above, we assume that $n_t \pi_{jt}$ is bounded below by a small constant $\underline{\varepsilon}_1 > 0$, then we can define

$$\underline{\iota}_u := \sup_{n, \pi: n\pi \geq \underline{\varepsilon}_1} \iota^*(n, n\pi) \quad \text{and} \quad \bar{\iota}_\ell := \inf_{n, \pi: n\pi \geq \underline{\varepsilon}_1} \iota^*(n, n\pi). \tag{11}$$

Furthermore, suppose that $\underline{\iota}_u$ and $\bar{\iota}_\ell$ are known and $\underline{\iota}_u < \infty$, $\bar{\iota}_\ell > 0$ for now, which we will discuss shortly below. Then, if we let ι_u and ι_ℓ be any finite number satisfying $\iota_u \geq \underline{\iota}_u$ and $0 < \iota_\ell \leq \bar{\iota}_\ell$, we will have

$$\begin{aligned} &E\bigl[\log\bigl((n_t s_{jt} + \iota_u)/n_t\bigr) - \log(\pi_{jt}) | z_{jt}\bigr] \geq 0 \quad \text{and} \\ &E\bigl[\log\bigl((n_t s_{jt} + \iota_\ell)/n_t\bigr) - \log(\pi_{jt}) | z_{jt}\bigr] \leq 0. \end{aligned} \tag{12}$$

Combining this with the orthogonality condition $E[\xi_{jt}|z_{jt}] = 0$, we obtain a set of conditional moment inequalities:

$$\begin{aligned} &E\bigl[\log(n_t s_{jt} + \iota_u)/n_t) - \log(\pi_{0t}) - x_{jt}'\beta_0 | z_{jt}\bigr] \geq 0 \\ &E\bigl[\log(n_t s_{jt} + \iota_\ell)/n_t) - \log(\pi_{0t}) - x_{jt}'\beta_0 | z_{jt}\bigr] \leq 0. \end{aligned} \tag{13}$$

The piece $\log(\pi_{0t})$ is easy to estimate because π_{0t} is typically large (sufficiently distant from zero) in most empirical work. We can plug in s_{0t} or any modification $\tilde{s}_{0t}$ of s_{0t} for

[11] Here, we maintain the standard assumption that in each given market, consumers' choices are independent and identically distributed.

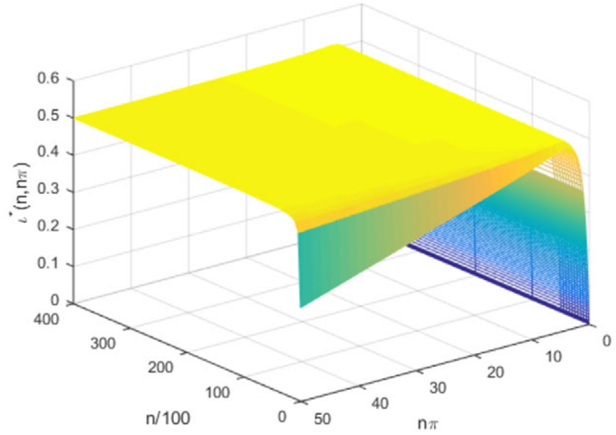

FIGURE 2. $\iota^*(n, n\pi)$ for a range of n and $n\pi$ values.

π_{0t}. As long as the modification is negligible relative to the estimation error in s_{0t}, standard arguments will imply $T^{-1}\sum_{t=1}^{T}[\log(\tilde{s}_{0t}) - \log(\pi_{0t})] = o_p(1)$. We specify $\tilde{s}_{0t}$ in the general case later. For the logit case, $\tilde{s}_{0t} = s_{0t}$ works just fine.

Now we discuss the choice of ι_ℓ and ι_u in greater detail. The first two questions we seek to answer are whether $\bar{\iota}_\ell$ is positive and $\underline{\iota}_u$ is finite, and whether we know them without the knowledge of the lower bound $\underline{\varepsilon}_1$ for $n_t\pi_{jt}$. The third question is how to choose ι_ℓ and ι_u given our answers to the first two questions.

We answer the first two questions by numerically obtaining $\iota^*(n, n\pi)$ for a large representative set of values of n and $n\pi$ and plot them in Figure 2.[12] The figure shows that $\iota^*(n, n\pi)$ varies smoothly with its two arguments, which gives us confidence that the supremum and the infimum from these discrete values are close to those of the function. Specifically, Figure 2 shows that $\underline{\iota}_u \approx 0.5$ and it is not affected by $\underline{\varepsilon}_1$. For $\bar{\iota}_\ell$, the figure shows that it approaches zero as $\underline{\varepsilon}_1$ approaches zero. Thus, without knowing $\underline{\varepsilon}_1$, we do not know $\bar{\iota}_\ell$. Nevertheless, the calculation that leads to Figure 2 also produces Table 2, which gives us an idea of how $\bar{\iota}_\ell$ changes with $\underline{\varepsilon}_1$. As the table shows, when $\underline{\varepsilon}_1$ is very small, $\bar{\iota}_\ell$ is well approximated by $\underline{\varepsilon}_1$.[13]

Given what we have learned about $\underline{\iota}_u$ and $\bar{\iota}_\ell$, we recommend choosing ι_u and ι_ℓ as follows. For ι_u, any $\iota_u > \underline{\iota}_u$ works in theory, but for better finite sample property, we recommend an ι_u a bit larger. In the Monte Carlo simulations, we find that $\iota_u = 2$ works well. Moreover, using $\iota_u = 2$ has an added benefit: it not only satisfies the theoretical requirement for the logit model, but also satisfies the requirement for nonlogit based models, as we will see in Section 4.2.

[12] We considered the values: $n \in \{100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 10{,}000, 20{,}000, 40{,}000\}$ and $n\pi \in \{0.0001 : 0.0001 : 0.01, 0.02 : .01 : 1, 1 : 0.1 : 50\}$, where the numbers between semicolons are the step sizes.

[13] Complete analytical investigation of the shape of $\iota^*(n, n\pi)$ is difficult due to the lack of analytical solution to expectations of the logarithm of binomial random variables. However, we provide some partial answers by analytically deriving the limit of $\iota^*(n, n\pi)$ as $n\pi$ approaches infinity and that as $n\pi$ approaches zero in Supplemental Appendix E. These limits are consistent with the numerical results reported in Figure 2 and Table 2.

TABLE 2. Computed $\bar{\iota}_\ell$ for various values of $\underline{\varepsilon}_1$.

$\underline{\varepsilon}_1$	n	$\bar{\iota}_\ell$
≥ 0.5	$\in [100, 40{,}000]$	≥ 0.250
0.1	$\in [100, 40{,}000]$	0.0776
0.01	$\in [100, 40{,}000]$	0.00955
0.001	$\in [100, 40{,}000]$	0.000993
0.0001	$\in [100, 40{,}000]$	0.0000998

For ι_ℓ, one can make a guess about how small $\underline{\varepsilon}_1$ can be based on institutional knowledge, and simply use an ι_ℓ that is smaller than this number. In practice, it sometimes is not difficult to make an educated guess of $\underline{\varepsilon}_1$ when you realize that $\underline{\varepsilon}_1$ is the lowest *number of units* that one expects a product to sell in a market. For example, if the market unit is week. and the product is a particular yogurt, the supermarket probably will not put it on the shelf if it is expected to sell less than one unit per 100 weeks. That gives us a lower bound $\underline{\varepsilon}_1 = 0.01$.

What if one makes a wrong guess at the lowest number of sales? Overguessing can cause violations of the moment inequalities (12), but fortunately, underguessing does *not*. Setting ι_ℓ at a value much lower than the actual $\bar{\iota}_\ell$ can guarantee the validity of (12). In our Monte Carlo and application, we in fact use an extremely low $\iota_\ell = 2^{-52}$ just to be on the safe side. As we see in the Monte Carlo and the empirical application, the estimates have good precision despite the extremely small ι_ℓ used.[14]

3.4 *Point estimation*

One can use any of the inference procedures for moment inequality models on (13), for example, Andrews and Shi (2013) and Cox and Shi (2019). Point identification is not required. On the other hand, point identification can greatly reduce the computational cost because inference without point identification generally requires costly test inversion. This is especially important for more complicated demand models than multinomial logit where even standard BLP estimation is computationally nontrivial.

In later sections, we discuss conditions that guarantee point identification. Under those conditions, the inequalities in (13) hold as equalities asymptotically on a set of z_{jt} values of positive measure, and ensure point identification in the same spirit as Kahn and Tamer (2009) in the context of endogenously censored regression models. To capture the identification information provided by those z_{jt} values, we consider a countable collection $\mathcal{G}$ of instrumental indicator functions $g : R^{d_z} \to \{0, 1\}$, where d_z is the dimension of z_{jt}. We adopt the collections of instrumental functions in Andrews and Shi (2013). Such collections are shown therein to preserve all the identification information in the

[14]We note that this is generally true if the point-identification condition in Section 5 below holds and n_t is large. But if the point-identification condition does not hold or n_t is too small, ι_ℓ can affect the precision of the inference. In that case, one should use institutional knowledge to carefully determine $\underline{\varepsilon}_1$—the lower bound for $n_t\pi_{jt}$, subsequently determine $\underline{\iota}_\ell$ according to Table 2, and choose $\iota_\ell = \underline{\iota}_\ell$.

conditional moment inequality model (13), and thus they preserve the point identification provided by the set of z_{jt} values at which the inequalities asymptotically hold as equalities, without that set of values being known. An example of $\mathcal{G}$ is given below.

We form the sample moments

$$\bar{m}_T^u(\beta, g) := (T\bar{J}_T)^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \big(\hat{\delta}_{jt}^u - x'_{jt}\beta\big) g(z_{jt}) \quad \text{and}$$

$$\bar{m}_T^\ell(\beta, g) := (T\bar{J}_T)^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \big(\hat{\delta}_{jt}^\ell - x'_{jt}\beta\big) g(z_{jt}),$$

where $\bar{J}_T = T^{-1} \sum_{t=1}^{T} J_t$, $\hat{\delta}_{jt}^u = \log((n_t s_{jt} + \iota_u)/n_t)$, and $\hat{\delta}_{jt}^\ell = \log((n_t s_{jt} + \iota_\ell)/n_t)$. These moments are used to form the criterion function:

$$\widehat{Q}_T(\beta) = \sum_{g \in \mathcal{G}} \mu(g) \big\{ \big[\bar{m}_T^u(\beta, g)\big]_-^2 + \big[\bar{m}_T^\ell(\beta, g)\big]_+^2 \big\}, \tag{14}$$

where $\mu(g) : \mathcal{G} \to [0, 1]$ is a probability mass function on $\mathcal{G}$, $[x]_- = \min\{0, x\}$ and $[x]_+ = \max\{0, x\}$. Finally, we define the estimator $\widehat{\beta}_T$ to be the minimizer of $\widehat{Q}_T(\beta)$:

$$\widehat{\beta}_T = \arg\min_{\beta \in B} \widehat{Q}_T(\beta), \tag{15}$$

where B is the parameter space of β. As we can see, computation of this estimator is on par with the standard GMM estimator for the multinomial logit model.

For $\mathcal{G}$, we divide the instrument vector z_{jt} into discrete instruments, $z_{d,jt}$, and continuous instruments $z_{c,jt}$. Without loss of generality, assume that $z_{c,jt}$ lies in $[0, 1]^{d_{z_c}}$.[15] Let the set $\mathcal{Z}_d$ be the discrete set of values that $z_{d,jt}$ can take. The set $\mathcal{G}$ is defined as

$$\begin{aligned} &\mathcal{G} = \big\{ g_{a,r,\zeta}(z_d, z_c) = 1\big((z'_c, z'_d)' \in C_{a,r,\zeta}\big) : C_{a,r,\zeta} \in \mathcal{C} \big\}, \quad \text{where} \\ &\mathcal{C} = \big\{ (\times_{u=1}^{d_{z_c}} \big((a_u - 1)/(2r), a_u/(2r)]\big) \times \{\zeta\} : a_u \in \{1, 2, \dots, 2r\}, \text{ for } u = 1, \dots, d_{z_c}, \\ &\qquad r = r_0, r_0 + 1, \dots, \text{ and } \zeta \in \mathcal{Z}_d \big\}. \end{aligned} \tag{16}$$

In practice, we truncate r at a finite value $\bar{r}_T$.[16] This does not affect the first-order asymptotic property of our estimator as long as $\bar{r}_T \to \infty$ as $T \to \infty$. For $\mu(\cdot)$, we use

$$\mu\big(\{g_{a,r,\zeta}\}\big) \propto (100 + r)^{-2} (2r)^{-d_{z_c}} K_d^{-1}, \tag{17}$$

where K_d is the number of elements in $\mathcal{Z}_d$. The same μ measure is used and works well in Andrews and Shi (2013).[17]

[15] If not, we can normalize it to lie in [0, 1] as suggested in Andrews and Shi (2013). For example, we can let $\tilde{z}_{c,jt} = F_{N(0,1)}(\widehat{\Sigma}_{z_c}^{-1/2} z_{c,jt})$, where $F_{N(0,1)}(\cdot)$ is the standard normal cdf and $\widehat{\Sigma}_{z_c}$ is the sample covariance matrix of $z_{c,jt}$, and use $\tilde{z}_{c,jt}$ in place of $z_{c,jt}$ to construct the instrumental functions.

[16] We shall show some simulation results in the Monte Carlo section that provides useful guidance on choosing $\bar{r}_T$ (and other ways of keeping the dimension of $\mathcal{G}$ manageable) in practice.

[17] Note that appropriate choices of $\mathcal{G}$ and μ are not unique. For other possible choices, see Andrews and Shi (2013).

4. The general model

Now we extend our discussion to the general differentiated product demand model and present our parameter estimator.

4.1 *Setup*

The specification of the general model is the same as the logit model except that the consumer level shock ϵ_{ijt} in $u_{ijt} = \delta_{jt} + \epsilon_{ijt} \equiv x'_{jt}\beta + \xi_{jt} + \epsilon_{ijt}$ is no longer type-I extreme value distribution. Instead, we assume that

$$\epsilon_{it} = (\epsilon_{i0t}, \dots, \epsilon_{iJ_t t}) \sim F\big(\cdot|x_t; \lambda\big), \tag{18}$$

where x_t stands for $(x'_{1t}, \dots, x'_{J_t t})'$, and $F(\cdot|x_t, \lambda)$ is a conditional cumulative distribution function known up to the finite-dimensional unknown parameter λ. By allowing x_t and an unknown parameter to enter the distribution of ϵ_{ijt}, this specification is general enough to encompass most models used in empirical work. In particular, it encompasses the random coefficient specifications $\epsilon_{ijt} = x'_{jt}(\beta_i - \beta) + \nu_{ijt}$, where β_i is a vector of random coefficients that follows a distribution (e.g., joint normal) known up to some unknown parameter, ν_{ijt} is the idiosyncratic taste shock.[18]

Given the specification, the unknown parameter in the general model is $\theta = (\beta', \lambda')'$. For clarity, we use $\theta_0 \equiv (\beta'_0, \lambda'_0)'$ to denote the true value of θ. Let $B \subseteq R^{d_\beta}$ denote the parameter space of β, and $\Lambda \subseteq R^{d_\lambda}$ the parameter space of λ. Let $\Theta = B \times \Lambda$ be the parameter space of θ.

In this model, the choice probability of each product is determined by

$$\pi_{jt} = \int 1\Big\{\delta_{jt} + \epsilon_j \geq \max_{j'=0,1,\dots,J_t} (\delta_{j't} + \epsilon_{j'})\Big\}\, dF(\epsilon_0, \epsilon_1, \dots, \epsilon_{J_t}|x_t, \lambda_0), \quad j = 0, 1, \dots, J_t. \tag{19}$$

Let $\pi_t = (\pi_{1t}, \dots, \pi_{J_t t})'$. This system is invertible under the connected substitute condition in Berry, Gandhi, and Haile (2013). In other words, we can define the inverse demand function $\delta_t(\pi_t, \lambda) := (\delta_{jt}(\pi_t, \lambda))_{j=1}^{J_t}$ as the solution to the system

$$\pi_{jt} = \int 1\Big\{\delta_{jt}(\pi_t, \lambda) + \epsilon_j \geq \max_{j'=0,1,\dots,J_t} \big(\delta_{j't}(\pi_t, \lambda) + \epsilon_{j'}\big)\Big\}\, dF(\epsilon_0, \epsilon_1, \dots, \epsilon_{J_t}|x_t, \lambda), \quad j = 1, \dots, J_t. \tag{20}$$

Inverting the demand system allows for the use of instrumental variables to identify θ based on the exclusion restriction:

$$E\big[\xi_{jt}|z_{jt}\big] = 0, \tag{21}$$

[18] Requiring $F(\cdot|x_t, \lambda)$ to be known up to a finite-dimensional parameter rules out the vertical model (see Berry and Pakes (2007)) because for the vertical model, ϵ_{ijt} is a function of the unobservable product characteristics (quality).

where z_{jt} is a vector of exogenous variables including exogenous components of x_{jt} and excluded instruments if there are any. This is because one can then obtain the following moment restriction:

$$E\big[\delta_{jt}(\pi_t, \lambda_0) - x'_{jt}\beta_0 | z_{jt}\big] = 0. \tag{22}$$

If π_t were observed, the parameter θ in the model would be identified under standard GMM identification conditions. However, as discussed in the logit case, π_t is not observed. Instead only a noisy measure $s_t := (s_{1t}, \dots, s_{J_t t})'$ is, and s_t frequently contains zero elements in many commonly used data sets. As in the logit case, $\delta_t(s_t, \lambda)$ is typically not well-defined when s_t contains zero elements, and thus simply substituting s_t for π_t in the moment conditions (22) is problematic.

4.2 *Bound construction*

Like in the logit case, we construct a pair of functions: $\hat{\delta}^u_{jt}(s_t, \lambda)$ and $\hat{\delta}^\ell_{jt}(s_t, \lambda)$, to form bounds for $\delta_{jt}(\pi_t, \lambda)$. The construction is based on the bounds for the logit case but adjusts for the different functional form:

$$\begin{aligned} \hat{\delta}^u_{jt}(s_t, \lambda) &= \log\big((n_t s_{jt} + \iota_u)/n_t\big) + \delta_{jt}(\tilde{s}_t, \lambda) - \log(\tilde{s}_{jt}), \\ \hat{\delta}^\ell_{jt}(s_t, \lambda) &= \log\big((n_t s_{jt} + \iota_\ell)/n_t\big) + \delta_{jt}(\tilde{s}_t, \lambda) - \log(\tilde{s}_{jt}), \end{aligned} \tag{23}$$

where ι_ℓ and ι_u are fixed numbers, and $\tilde{s}_t$ is a slight modification of s_t to take it off the boundary of the probability simplex. We will require that the modification of $\tilde{s}_{jt}$ to s_{jt} is small so that $\|\tilde{s}_t - s_t\| = O_p(1/n_t)$. For example, $\tilde{s}_{jt} = s_{jt} + 1/n_t$ (when J_t is bounded) or $\tilde{s}_{jt} = s_{jt} + 1/(n_t J_t)$ (when J_t is unrestricted) for $j = 1, \dots, J_t$.[19] The conditions for ι_u and ι_ℓ are specified in Assumptions 1(b) and 2(c) below. We recommend $\iota_u = 2$ and $\iota_\ell = 2^{-52}$ and use these choices in the Monte Carlo and application, as discussed in Section 3.3.

To see why the construction in (23) may be valid and what requirements we may need on ι_u and ι_ℓ, consider the upper bound for example:

$$\begin{aligned} \hat{\delta}^u_{jt}(s_t, \lambda) - \delta_{jt}(\pi_{jt}, \lambda) = {} & \big[\log\big((n_t s_{jt} + \iota_u)/n_t\big) - \log(\pi_{jt})\big] \\ & + \big[\big(\delta_{jt}(\tilde{s}_t, \lambda) - \log(\tilde{s}_{jt})\big) - \big(\delta_{jt}(\pi_t, \lambda) - \log(\pi_{jt})\big)\big]. \end{aligned}$$

We already know from the logit case that the first summand is nonnegative in expectation conditional on π_{jt} as long as $\iota_u \geq \underline{\iota}_u$ for $\underline{\iota}_u$ defined in equation (11). It is then clear that the bound $\hat{\delta}^u_{jt}(s_t, \lambda)$ will be asymptotically valid if either (i) the conditional expectation of the second summand is asymptotically negligible, or (ii) the conditional expectation of the sum of the two is non-negative. Next, we show that the first case applies to logit-based models, while the second case applies to models where the idiosyncratic error has a thinner tail than the logistic distribution, for example, normal distributions.

[19]We note that this implies $\tilde{s}_{0t} = s_{0t} - J_t/n_t$ or $\tilde{s}_{0t} = s_{0t} - 1/n_t$. This in principle could be less than or equal to zero. But in typical data sets, this is not an issue because s_{0t} is much larger than J_t/n_t. It is not an issue asymptotically as we will assume that π_{0t}—the share of the outside good—is bounded away from zero.

Case 1. When $\delta_{jt}(\cdot, \lambda) - \log(\cdot_j)$ is uniformly continuous.

Let $\Delta^0_{J_t}$ denote the subset of $\{\pi \in (0, 1)^{J_t} : \mathbf{1}'_{J_t}\pi < 1\}$ that π_t can take value in. Let $\Delta^c_{J_t}$ denote an c-expansion of $\Delta^0_{J_t}$, that is, $\Delta^c_{J_t} = \{\pi \in (0, 1)^{J_t} : \pi'\mathbf{1}_{J_t} < 1, \min_{p\in\Delta^0_{J_t}} \|p-\pi\|_f \le c\}$ for $c > 0$, where $\|p - \pi\|_f = \sqrt{\|p - \pi\|^2 + (\mathbf{1}'(p - \pi))^2}$. Note that the metric $\|\cdot\|_f$ takes into account the difference for the outside share, while the Euclidean norm on $\{\pi \in (0, 1)^{J_t} : \mathbf{1}'_{J_t}\pi < 1\}$ only considers the shares for the inside goods.

Define the function $\check{\delta}_t(\cdot, \lambda) = (\check{\delta}_{1t}(\cdot, \lambda), \dots, \check{\delta}_{J_t t}(\cdot, \lambda))' : \Delta^c_{J_t} \to R^{J_t}$ where

$$\check{\delta}_{jt}(\pi, \lambda) := \delta_{jt}(\pi, \lambda) - \log(\pi_j).$$

Since $\Delta^c_{J_t}$ may contain points arbitrarily close to the boundary of the probability simplex, in general neither $\delta_{jt}(\cdot, \lambda)$ nor $\log(\cdot_j)$ is uniformly continuous on $\Delta^c_{J_t}$. Thus, neither $\delta_{jt}(\tilde{s}_t, \lambda) - \delta_{jt}(\pi_t, \lambda)$ nor $\log(\tilde{s}_{jt}) - \log(\pi_{jt})$ may converge to zero as $n_t \to \infty$ and $\pi_{jt} \to 0$ even if $\tilde{s}_t$ is the most efficient consistent estimate of π_t. However, in many models used in empirical work, the logit inverse demand $(\log(\pi_j) - \log(\pi_0))$ is a good first-order approximation of $\delta_{jt}(\pi, \lambda)$ when π_j is close to zero and this first-order term is the entire reason that the inverse demand is not uniformly continuous. For such models, the following assumption is reasonable.

Assumption 1. (a) *For some $c > 0$,*

$$\max_{t=1,\dots,T; j=1,\dots,J_t} \sup_{\pi,\tilde{\pi}\in\Delta^c_{J_t}:\pi\neq\tilde{\pi}} \sup_{\lambda\in\Lambda} \frac{|\check{\delta}_{jt}(\tilde{\pi}, \lambda) - \check{\delta}_{jt}(\pi, \lambda)|}{\|\tilde{\pi} - \pi\|_f \sqrt{J_t}} \le O(1).$$

(b) $0 < \iota_\ell \le \bar{\iota}_\ell$ *and* $\underline{\iota}_u \le \iota_u < \infty$, *where* $\bar{\iota}_\ell$ *and* $\underline{\iota}_u$ *are defined in equation* (11), *and* $\sup_{t=1,\dots,T} n_t\|\tilde{s}_t - s_t\|_f = O_p(1)$.

Now we give two examples where Assumption 1(a) is satisfied.

Example 4.1. Nested logit. The inverse demand of the nested logit model can be written as $\delta_{jt}(\pi_t, \lambda) = \log(\pi_{jt}/\pi_{0t}) - \lambda \log(\pi_{gt}/\pi_{0t})$ where π_{gt} is the aggregate share of all the products in the nest (nest g) that j is in. In this case, $\check{\delta}_{jt}(\pi_t, \lambda) = (\lambda - 1)\log \pi_{0t} - \lambda \log(\pi_{gt})$. Assumption 1(a) is satisfied if $\Delta^0_{J_t} = \{\pi \in (0, 1)^{J_t} : 1 - \mathbf{1}'_{J_t}\pi > \underline{\varepsilon}_0, \pi_{gt} > \underline{\varepsilon}_0$ for all nests $g\}$. In fact, Assumption 1(a) holds without the $\sqrt{J_t}$, which is a stronger version of the assumption. The requirement that π_{0t} and π_{gt} are bounded away from zero is reasonable for data sets in which neither the outside good nor any of the nests have zero shares.

Example 4.2. Random coefficient logit. For the random coefficient logit model, $\delta_{jt}(\pi_t; \lambda)$ is the solution to the following equation system:

$$\pi_{jt} = \exp(\delta_{jt}) \int \frac{\exp(w'_{jt}v)}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + w'_{kt}v)} dF(v; \lambda), \quad j = 1, \dots, J_t,$$

where w_{jt} is a vector of covariates with random coefficients, and $F(\cdot;\lambda)$ is the distribution of the random coefficient known up to the unknown parameter λ. Using the definition of $\check{\delta}_{jt}$ above, we can write

$$\exp\bigl(-\check{\delta}_{jt}(\pi_t;\lambda)\bigr) = \int \frac{\exp\bigl(w_{jt}'v\bigr)}{1+\sum_{k=1}^{J_t}\exp\bigl(\check{\delta}_{kt}(\pi_t;\lambda)+w_{kt}'v\bigr)\pi_{kt}}\,dF(v;\lambda). \tag{24}$$

Assume that $\|w_{jt}\|$ is bounded by $\overline{w}$ and $0<\sup_{w:\|w\|\leq\overline{w}}\int\exp(w'v)\,dF(v;\lambda)<\infty$. We can already see that $\check{\delta}_{jt}(\pi_t;\lambda)$ is bounded away from $-\infty$ when $\pi_{jt}\to 0$ (in which case, $\delta_{jt}(\pi_t;\lambda)\to-\infty$). With additional algebra, we can show that $\partial\check{\delta}_{jt}(\pi_t;\lambda)/\partial\pi_t$ is bounded, which essentially guarantees Assumption 1(a). The details are given in Supplemental Appendix D.

Under Assumption 1(a), the requirement for ι_u and ι_ℓ are the same as in the logit case, which is formally stated in Assumption 1(b).

Case 2: when $\delta_{jt}(\cdot,\lambda)-\log(\cdot_j)$ *is not uniformly continuous.*

In some models used in empirical work, Assumption 1 can fail to hold. For example, if the model is a simple probit with $J_t=1$, $\delta_t(\pi)=\Phi^{-1}(\pi)$, where Φ^{-1} is the inverse of the standard normal cdf. In this case, $\check{\delta}_t(\pi)=\delta_t(\pi)-\log(\pi)=\Phi^{-1}(\pi)-\log\pi$. This function approaches $+\infty$ when $\pi\to 0$, and has arbitrarily large slope near zero. For such cases, an alternative assumption may be reasonable and this is given in the following assumption.

Assumption 2. (a) $\max_{j,t}E[\check{\delta}_{jt}(\tilde{s}_t,\lambda_0)-\check{\delta}_{jt}(\pi_t,\lambda_0)|\pi_t,z_t]\leq 0$,

(b) $\min_{j,t}E[\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0)|\pi_t,z_t]\geq 0$, *and*

(c) *for* $j=1,\dots,J_t$, $\tilde{s}_{jt}=s_{jt}+1/n_t$, $0<\iota_\ell\leq\overline{\iota}_\ell$ *where* $\overline{\iota}_\ell$ *is defined in equation* (11), *and* $1<\iota_u<\infty$.

(d) $\sup_j\sup_{\pi:\pi_j\geq(\underline{\varepsilon}_1\wedge 1)/n_t}|\delta_{jt}(\pi,\lambda_0)|\leq C_0\log(n_t)$ *for a constant* $C_0>0$ *for all* t.

Example 4.3. Binary probit. For the binary probit model, we verified numerically that parts (a)–(b) hold given part (c), even though we do not have a theoretical proof. The intuition is that $\Phi^{-1}(\pi)$ decreases slower than $\log(\pi)$ as $\pi\to 0$. Thus, smaller modification ι is needed for $E[\Phi^{-1}(s_{jt}+\iota/n_t)|\pi_t,z_t]$ to exceed $E[\Phi^{-1}(\pi_{jt})|\pi_t,z_t]$ than for $E[\log(s_{jt}+\iota/n_t)|\pi_t,z_t]$ to exceed $E[\log(\pi_{jt})|\pi_t,z_t]$. And $\iota=1$ is sufficient for the latter, as we discussed in the logit case. Part (d) holds simply because of the shape of $\Phi^{-1}(\cdot)$, which increases slower than $\log(\cdot)$ as the argument decreases to zero.

The following lemma shows that the bounds constructed in (23) are asymptotically valid.

Lemma 1. *Suppose that* $\min_{t=1,\dots,T}n_t\to\infty$ *as* $T\to\infty$, *that* $n_t\pi_{jt}\geq\underline{\varepsilon}_1$ *for* $j=1,\dots,J_t$ *and* $\underline{\varepsilon}_1$ *being the positive number used in* (11), *and that* $E[\xi_{jt}|z_{jt}]=0$ *if either Assumption*

1 *or Assumption* 2(a)–(c) *holds. Then there exist random variables* e^u_{jt} *and* e^ℓ_{jt} *such that* $\sup_{j=1,\dots,J_t;t=1,\dots,T} \frac{n_t^{1/2}}{T^{1/4}J_t^{1/2}}|e^y_{jt}| = O_p(1)$ *for* $y = u, \ell$, *and*

$$E\left[\hat{\delta}^u_{jt}(s_t, \lambda_0) - x_{jt}\beta_0 + e^u_{jt}|z_{jt}\right] \geq 0,$$
$$E\left[\hat{\delta}^\ell_{jt}(s_t, \lambda_0) - x_{jt}\beta_0 + e^\ell_{jt}|z_{jt}\right] \leq 0. \tag{25}$$

The moment inequalities (25) can be taken to the data since the term e^y_{jt} $(y = u, \ell)$ is ignorable provided that n_t increases at a faster rate than $T^{1/2}J_t$. Also, note that for the multinomial logit and the nested logit case, the lemma holds without the $J_t^{1/2}$ in the denominator because Assumption 1(a) holds without the $J_t^{1/2}$ in $\sup_{j=1,\dots,J_t;t=1,\dots,T} \frac{n_t^{1/2}}{T^{1/4}J_t^{1/2}}|e^y_{jt}| = O_p(1)$. Thus, for these models n_t only needs to increase faster than $T^{1/2}$.

4.3 *Point estimation*

Like in the logit case, one can apply any of the inference procedures for moment inequality models on (25). Yet point identification can greatly simplify computation. Point-identification conditions are given in Section 5. Under those conditions, the inequalities in (25) hold as equalities asymptotically on a set of z_{jt} values with positive measure.

We define the point estimator analogous to the logit case:

$$\widehat{\theta}_T := \left(\widehat{\beta}'_T, \widehat{\lambda}'_T\right)' = \arg\min_{\theta\in\Theta} \widehat{Q}_T(\theta), \tag{26}$$

where

$$\widehat{Q}_T(\theta) = \sum_{g\in\mathcal{G}} \mu(g)\left\{\left[\bar{m}^u_T(\theta, g)\right]^2_- + \left[\bar{m}^\ell_T(\theta, g)\right]^2_+\right\}, \quad \text{with} \tag{27}$$

$$\bar{m}^u_T(\theta, g) := (T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left(\hat{\delta}^u_{jt}(s_t, \lambda) - x'_{jt}\beta\right)g(z_{jt}) \quad \text{and}$$

$$\bar{m}^\ell_T(\theta, g) := (T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left(\hat{\delta}^\ell_{jt}(s_t, \lambda) - x'_{jt}\beta\right)g(z_{jt}),$$

where $\mu(g) : \mathcal{G} \to [0, 1]$ is a probability mass function on $\mathcal{G}$ and $\mathcal{G}$ is a collection of instrumental functions. Both $\mathcal{G}$ and $\mu(\cdot)$ have been given in Section 3.4.

5. Point-identification condition

The point-identification condition is motivated by the power law feature of the data demonstrated in Section 2. The feature indicates a coexistence of a few dominant products with thick demand and a large number of fringe products with thin demand. For a dominant product j in market t, π_{jt} is large and $\log(s_{jt} + \iota_u/n_t)$ and $\log(s_{jt} + \iota_\ell/n_t)$ are close to each other and close to $\log(\pi_{jt})$ at large n_t. If certain values of the exogenous variables z_{jt} predict such π_{jt}'s, then at those z_{jt} values, the conditional moment inequalities in (25) hold as equalities asymptotically. These equalities may yield point identification by standard identification arguments for BLP moment conditions.

Formally, let $\mathcal{Z}_0$ stand for the set of values of z_{jt} that predict dominant products (those with choice probabilities that do not approach zero). We can state the assumption as follows.

Assumption 3. *There exists a fixed constant* $\underline{\varepsilon}_0 \in (0, 1)$ *and a set* $\mathcal{Z}_0 \subseteq \text{supp}(z_{jt})$ *such that* $\inf_{j,t,T} \Pr(z_{jt} \in \mathcal{Z}_0) > 0$, *such that* $\Pr(\pi_{jt} \geq \underline{\varepsilon}_0 | z_{jt} \in \mathcal{Z}_0) = 1$ *for all* j, t.

Below we give three stylized demand–supply models that could give rise to the dominant products and discuss what the dominant product predictors are in each case. For now, it is important to note that z_{jt} includes both the exogenous covariates in the demand model and excluded instruments (if there are any). Often in practice, it is brand or UPC dummies that predict dominant status, which usually are also included exogenous covariates.

We state the lemma that shows that the bounds collapse on $\mathcal{Z}_0$ under Assumption 3.

Lemma 2. *Suppose that* $\min_{t=1,\dots,T} n_t^2/T \to \infty$ *as* $T \to \infty$, *and that Assumption* 3 *holds. Then we have*

$$\sup_{j=1,\dots,J_t, t=1,\dots,T} \sup_{\lambda \in \Lambda} n_t \big|\hat{\delta}^u_{jt}(s_t, \lambda) - \hat{\delta}^\ell_{jt}(s_t, \lambda)\big| 1\{z_{jt} \in \mathcal{Z}_0\} = O_p(1). \tag{28}$$

Remark. When the bounds collapse, the moment inequalities (25) hold as equalities on $\mathcal{Z}_0$ asymptotically. Then the standard (point) identification considerations for BLP models apply here, except that attention is restricted on $\mathcal{Z}_0$. In general, if the instruments shift price and sales sufficiently for the dominant products, the model is point identified.

Remark. Note that neither $\mathcal{Z}_0$ or $\underline{\varepsilon}_0$ need to be known in order to use our estimator. This is an advantage of the moment inequality approach compared to an alternative approach that preselects products that never experience zeroes. The key to this is the Andrews and Shi (2013) type instrumental functions that ensure that asymptotically all of the information in the conditional moment inequalities (25) are preserved in forming the unconditional moments. That will guarantee that the point-identification information provided by $\mathcal{Z}_0$ is preserved as well, even though $\mathcal{Z}_0$ is unknown.

Next, we discuss how the dominant products may come into being. Such products or firms have been a subject of interest since the early days of industrial organization. They have been studied under the name of *incumbents*, *leaders* as well as as dominant products/firms (see, e.g., Markham (1951), Chapter 8 of Tirole (1988), Gowrisankaran and Holmes (2004), and Shimomura and Thisse (2012)). They are the ones that enjoy a large market share and earn a positive profit despite that there are free entry and an unlimited number of potential entrants. The literature does not agree on how they achieve their dominant status. Simple explanations include (a) the dominant products are less substitutable with the fringe products than the fringe products among themselves, (b) the dominant products are much more appealing on average possibly due to brand loyalty or technological innovations, (c) the dominant products are provided with significantly lower cost possibly due to technology advances. In all explanations, a key is that the dominant products have features that are not easily replicable, so there is no free entry of products with those features. We illustrate each using a stylized example now.[20] In the examples, we ignore the t subscript for notational ease.

Example 5.1. Consider a nested logit model with three nests: $\{0\}$, $\mathcal{J}_0$, $\mathcal{J}_1$, where $\mathcal{J}_0 \cup \mathcal{J}_1 = \{1, \dots, J\}$. Let J_0 and J_1 denote the number of elements in $\mathcal{J}_0$ and $\mathcal{J}_1$, respectively, and suppose that J_0 is fixed as n grows but J_1 grows proportionally to n, say $J_1 = cn$. Let $\pi_{\mathcal{J}_\ell}$ stand for the probability that a product in $\mathcal{J}_\ell$ is chosen, for $\ell = 0, 1$. Consider a nested logit model that yields

$$\begin{aligned} \frac{\pi_j}{\pi_{\mathcal{J}_\ell}} &= \frac{\exp(\delta_j)}{\sum\limits_{j' \in \mathcal{J}_\ell} \exp(\delta_{j'})} \quad \text{for } j \in \mathcal{J}_\ell, \\ \pi_{\mathcal{J}_\ell} &= \frac{\exp\big(\lambda\big(\mathcal{I}(\mathcal{J}_\ell) - \log(J_\ell)\big)\big)}{1 + \exp\big(\lambda\big(\mathcal{I}(\mathcal{J}_0) - \log(J_0)\big)\big) + \exp\big(\lambda\big(\mathcal{I}(\mathcal{J}_1) - \log(J_1)\big)\big)}, \\ &\quad \text{for } \ell = 0, 1, \end{aligned} \tag{29}$$

where $\mathcal{I}(\mathcal{J}_\ell) = \log(\sum_{j \in \mathcal{J}_\ell} \exp(\delta_j))$ and λ is a parameter. Suppose that $\delta_j : j = 1, \dots, J$ are bounded between $\underline{\delta}$ and $\overline{\delta}$. Then it is easy to verify that $\mathcal{I}(\mathcal{J}_\ell) - \log(J_\ell)$ is also bounded between $\underline{\delta}$ and $\overline{\delta}$. Thus, $\pi_{\mathcal{J}_\ell} \in [\frac{\exp(\lambda\underline{\delta}))}{1+\exp(\lambda\underline{\delta})+\exp(\lambda\overline{\delta})}, \frac{\exp(\lambda\overline{\delta}))}{1+\exp(\lambda\underline{\delta})+\exp(\lambda\overline{\delta})}]$, and $\frac{\pi_j}{\pi_{\mathcal{J}_\ell}} \in J_\ell^{-1}[\exp(\underline{\delta} - \overline{\delta}), \exp(\overline{\delta} - \underline{\delta})]$. Then we have

$$\begin{aligned} \pi_j &\ge J_0^{-1} \exp(\underline{\delta} - \overline{\delta} + \lambda\underline{\delta})/\big(1 + \exp(\lambda\underline{\delta}) + \exp(\lambda\overline{\delta})\big) \quad \text{for } j \in \mathcal{J}_0, \\ n\pi_j &\ge c^{-1} \exp(\underline{\delta} - \overline{\delta} + \lambda\underline{\delta})/\big(1 + \exp(\lambda\underline{\delta}) + \exp(\lambda\overline{\delta})\big) \quad \text{for } j \in \mathcal{J}_1. \end{aligned} \tag{30}$$

That is, products in nest $\mathcal{J}_0$ are dominant products satisfying $\pi_j \ge \underline{\varepsilon}_0$ and those in nest $\mathcal{J}_1$ are fringe products satisfying $n\pi_j \ge \underline{\varepsilon}_1$ for $\underline{\varepsilon}_0 = cJ_0^{-1}\underline{\varepsilon}_1 = J_0^{-1} \exp(\underline{\delta} - \overline{\delta} + \lambda\underline{\delta})/(1 + \exp(\lambda\underline{\delta}) + \exp(\lambda\overline{\delta})))$ for $j \in \mathcal{J}_1$. Assumption 3 is satisfied if $1\{j \in \mathcal{J}_0\}$ is part of z_{jt}.

In this example, the number of fringe products is proportional to n. This appears arbitrary, but can in fact be a natural result of the zero-profit condition under free entry into nest $\mathcal{J}_1$ (see the discussion at the end of Section 3.1). The dominant products are dominant because they are protected from the competition of the fringe products by the substitution pattern in product demand and barrier to entry into nest $\mathcal{J}_0$.

[20] As we can see, in each of the examples, the dominant status indicator is a discrete random variable. It is possible to conjure up a continuous dominant status indicator, but its support would need to have a discontinuity to separate the dominant and fringe products, a feature that could be difficult to justify in practice.

Example 5.2. Consider a multinomial logit model for simplicity. Normalize $\delta_{0t}=0$. Let

$$\delta_j = -\alpha p_j + \sum_{k=1}^{J} \beta_k \mathrm{UPC}_{kj} + \xi_j, \tag{31}$$

where p_j is the price, UPC_{kj}'s are UPC dummies ($\mathrm{UPC}_{kj} = 1\{k=j\}$), $\beta_j = b_j$ for $j \in \mathcal{J}_0$, and $\beta_j = -\log(n) + b_j$ for $j \notin \mathcal{J}_0$ for a subset $\mathcal{J}_0$ of $\{1, \dots, J\}$, and b_j are bounded constants. Let J be fixed. Let p_j and ξ_j be bounded. Then for $j \notin \mathcal{J}_0$,

$$\pi_j = \frac{\exp(-\alpha p_j + b_j + \xi_j)/n}{1 + \sum_{k=1}^{J} \exp(-\alpha p_k + \beta_k + \xi_k)} \geq n^{-1} \frac{\exp(-\alpha \overline{p} + \underline{b} + \underline{\xi})}{1 + J \exp(-\alpha \underline{p} + \overline{b} + \overline{\xi})}, \tag{32}$$

where $\underline{p}, \underline{b}, \underline{\xi}$ are the lower bounds of p_j, b_j, ξ_j, and $\overline{p}, \overline{b}, \overline{\xi}$ are the upper bounds. Let $\underline{\varepsilon}_1 = \frac{\exp(\alpha \underline{p} + \underline{b} + \underline{\xi})}{1 + J \exp(\alpha \overline{p} + \overline{b} + \overline{\xi})}$. Then this shows that $n\pi_j \geq \underline{\varepsilon}_1$. For $j \in \mathcal{J}_0$,

$$\pi_j = \frac{\exp(-\alpha p_j + b_j + \xi_j)}{1 + \sum_{k=1}^{J} \exp(-\alpha p_k + \beta_k + \xi_k)} \geq \frac{\exp(-\alpha \overline{p} + \underline{b} + \underline{\xi})}{1 + J \exp(-\alpha \underline{p} + \overline{b} + \overline{\xi})}. \tag{33}$$

Let $\underline{\varepsilon}_0 = \underline{\varepsilon}_1$. Then this shows that Assumption 3 holds if the UPC dummies are used as part of z_{jt}.[21]

In this example, the mean utility of the fringe products depends on n. This can be a natural result of the zero-profit condition under free entry (see the discussion at the end of Section 3.1): only fringe products with such mean utilities self-select into the market.

Example 5.3. Consider a multinomial logit model again. Let $\delta_j = -\alpha p_j + \xi_j$. Let there be constant marginal cost $c_j = \alpha^{-1} \log(n) z_j + c_{0j}$, where z_j is a dummy variable and c_{0j} is a bounded constant. Suppose for simplicity that the products are supplied by single-product firms maximizing profit. Then it is easy to see that the optimal price is

$$p_j = c_j + \frac{1}{\alpha(1-\pi_j)} \tag{34}$$

Let J be fixed and ξ_j be bounded. Then, for every $j = 1, \dots, J$,

$$\begin{aligned} \pi_j &= \frac{\exp\bigl(-\log(n) z_j - \alpha c_{0j} - (1-\pi_j)^{-1}\bigr)}{1 + \sum_{k=1}^{J} \exp\bigl(-\log(n) z_k - \alpha c_{0k} - (1-\pi_k)^{-1}\bigr)} \\ &\leq \exp(-\alpha \underline{c} - 1), \end{aligned} \tag{35}$$

[21] Note that letting α be a random coefficient or adding other covariates would not change the essence of the example.

where $\underline{c}$ is the lower bound for c_{0j}. Let $\overline{\pi} = \exp(-\alpha\underline{c} - 1)$. Then for j's with $z_j = 1$, we have

$$\pi_j \geq n^{-1} \frac{\exp\bigl(-\alpha\overline{c} - (1 - \overline{\pi})^{-1}\bigr)}{1 + \sum_{k=1}^{J} \exp(-\alpha\underline{c} - 1)}, \tag{36}$$

where $\overline{c}$ is the upper bound for c_{0j}. Let $\underline{\varepsilon}_1 = \frac{\exp(-\alpha\overline{c} - (1-\overline{\pi})^{-1})}{1+\sum_{k=1}^{J} \exp(-\alpha\underline{c}-1)}$. Then this shows that $n\pi_j \geq \underline{\varepsilon}_1$. Similarly, we can show that for j's with $z_j = 0$, $\pi_j \geq \underline{\varepsilon}_0 := \underline{\varepsilon}_1$, verifying Assumption 3.

In this example, the cost of the fringe products depends on n. This can be a natural result of the zero-profit condition under free entry (see the discussion at the end of Section 3.1): only fringe products with such costs self-select into the market.

As we see above, the point-identification assumption is natural in many situations with dominant products. Nevertheless, in settings where these assumptions are questionable, we can still use (25) as a basis for partial identification and inference. For example, one can use the method developed in Andrews and Shi (2013) to construct a joint confidence set for the full vector θ_0. This confidence set is constructed by inverting an Anderson–Rubin test: $CS = \{\theta : T(\theta) \leq c(\theta)\}$ for some test statistic $T(\theta)$ and critical value $c(\theta)$. Computing this set amounts to computing the 0-level set of the function $T(\theta) - c(\theta)$, where $c(\theta)$ typically is simulated quantiles, and thus a nonsmooth function of θ. A new approach that is computationally less burdensome when β is high dimensional is proposed in Gandhi, Lu, and Shi (2013), which also includes Monte Carlo simulations and empirical results using the profiling approach under partial identification. When the linear coefficients of the control variables are nuisance parameters, one can also use the approach in Cox and Shi (2019) for inference to further reduce computational burden.

6. Consistency

In this section, we establish the consistency of the point estimator defined in (26). We need additional assumptions.

The first set of assumptions formalize the model and the data environment. They are similar to those in Berry, Linton, and Pakes (2004) and Freyberger (2015).

Assumption 4. (a) *The equation system* (20) *uniquely defines* $\delta_t(\pi_t, \lambda)$ *for all* t, *all* $\pi_t \in \{\pi \in (0,1)^{J_t} : \mathbf{1}'_{J_t}\pi < 1\}$, *and all* $\lambda \in \Lambda$.

(b) *In each market, consumers' preferences* $(\epsilon_{ijt})_{j=1}^{J_t}$ *are i.i.d. draws from the known distribution* $F(\cdot|x_t; \lambda_0)$ with unknown parameter $\lambda_0 \in \Lambda$. *Consumer choice is determined by* (19).

(c) *The moment condition* (22) *holds.*

(d) $(x_t, s_t, z_t)_{t=1}^{T}$ *are independent across markets.*

(e) *There exists a constant* M *such that* $E[\xi_{jt}^{2+c}] < M$ *for all* $j = 1, \ldots, J_t$, *all* $t = 1, \ldots, T$, *and all* T *for some* $c > 0$.

(f) $\sup_{t=1,\dots,T} n_t\|\tilde{s}_t - s_t\|_f = O_p(1)$ *as* $T \to \infty$.

(g) $\frac{\underline{n}_T}{J_T^{\max}\sqrt{T}} \to \infty$ *and* $\frac{\log(\overline{n}_T)}{\sqrt{T}} \to 0$ *where* $\underline{n}_T = \min_{t=1,\dots,T} n_t$, $J_T^{\max} = \max_{t=1,\dots,T} J_t$, *and* $\overline{n}_T = \max_{t=1,\dots,T} n_t$.

Remark. Part (g) requires that n_t be not too small and not too big. The not-too-big part may be surprising because larger n_t is typically considered a good thing. Here, larger n_t is not purely a good thing because we allow the lowest π_{jt} to be inversely related to n_t.[22] In this framework, larger n_t also implies lower minimum π_{jt}, which *increases* the difficulty in bounding $\log(\pi_{jt})$. Also, note that the $J_T^{\max}$ in part (g) is not needed for multinomial logit and nested logit models for the reason discussed in the paragraph below Lemma 1.

The next assumption formalizes the lower bound for choice probabilities for the outside and the fringe products. These bounds have been discussed in detail in Sections 3 and 5.

Assumption 5. (a) $\pi_{0t} > \underline{\varepsilon}_0$ *for all* t.

(b) $\pi_{jt} > \underline{\varepsilon}_1/n_t$ *for all* j, t.

Next, we impose a Lipschitz continuity assumption on $\delta_{jt}(\pi, \lambda)$ in π on the part of the π space for the dominant products.

Assumption 6.

$$\sup_{t=1,\dots,T}\ \sup_{j=1,\dots,J_t}\ \sup_{\lambda\in\Lambda}\ \sup_{\pi,\tilde{\pi}\in\Delta_{J_t}^{\underline{\varepsilon}_0/2}:\pi\neq\tilde{\pi},\pi_j,\tilde{\pi}_j\geq\underline{\varepsilon}_0/2} \frac{|\delta_{jt}(\tilde{\pi}, \lambda) - \delta_{jt}(\pi, \lambda)|}{\|\tilde{\pi} - \pi\|_f\sqrt{J_t}} = O(1).$$

Remark. Assumption 6 is a commonly accepted assumption when all products are dominant products (ref. Freyberger (2015)). The stronger version of this assumption without $\sqrt{J_t}$ on the denominator holds for multinomial logit models: $\delta_{jt}(\pi, \lambda) = \log(\pi_j) - \log(\pi_0)$ because the logarithm function is uniformly continuous on the interval $[\underline{\varepsilon}_0/2, \infty)$. This argument combined with Assumption 1 (a) implies Assumption 6 for models satisfying Assumption 1. The same argument as that for the multinomial logit also works for the binary probit model.

Finally, we strengthen the point-identification condition to ensure consistency. Define

$$\mathcal{G}_0 = \{g \in \mathcal{G} : g(z) = 0 \text{ for } z \notin \mathcal{Z}_0\}. \tag{37}$$

This is the set of instrumental functions that captures the identification information provided by the dominant products. Note that the dominant status predictor(s) in z_{jt} often is (are) brand/UPC dummy(ies); thus, elements in $\mathcal{G}_0$ are often those dummies interacted with dummies created for other elements of z_{jt} in the Andrews and Shi (2013)

[22] Recall from Section 3 that this is done to rationalize the zeroes in the data.

style (described in Section 3.4). It is also worth noting that one does not need to know $\mathcal{G}_0$ but only need to know that $\mathcal{G}$ contains such a $\mathcal{G}_0$, the latter guaranteed by Assumption 3 and the Andrews and Shi style $\mathcal{G}$.

Let

$$\begin{aligned} \bar{m}_T(\theta, g) &= \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \bigl(\delta_{jt}(\pi_t, \lambda) - x'_{jt}\beta\bigr) g(z_{jt}), \\ \widehat{Q}^*_T(\theta) &= \sum_{g \in \mathcal{G}_0} \mu(g) \bar{m}_T(\theta, g)^2. \end{aligned} \tag{38}$$

The moments $\bar{m}_T(\theta, g)$ is infeasible because π_{jt} is not observed. But we will be able to show that they are close to $\bar{m}^u_T(\theta, g)$ and $\bar{m}^\ell_T(\theta, g)$ for $g \in \mathcal{G}_0$. The criterion function $\widehat{Q}^*_T(\theta)$ aggregates the infeasible moments for the dominant products. The assumption below is the additional identification condition.

Assumption 7. *For any $c > 0$, there exists $C(c) > 0$ such that*

$$\lim_{T\to\infty} \Pr\Bigl(\inf_{\theta \in \Theta : \|\theta^s - \theta^s_0\| > c} \widehat{Q}^*_T(\theta) > C(c)\Bigr) = 1,$$

where θ^s is a subvector of θ and θ^s_0 is its true value.

Remark. This assumption ensures that the dominant products provide enough restriction to point identify the parameter θ^s. Only a subvector of θ is considered in this assumption because we want to allow (but not require) specifications with product fixed effects. The fixed effects for the fringe products are clearly not identified since the data do not contain sufficiently precise information about their inverse demand. In that case, θ^s will only contain the common parameters and the fixed effects of the dominant products. Moreover, the assumption requires that the instrumental functions in $\mathcal{G}_0$ are able to capture the variation of the moments over $z_{jt} \in \mathcal{Z}_0$. This in general requires that $E[\hat{\delta}^u_{jt}(s_t, \lambda) - x_{jt}\beta | z_{jt} = z]$ and $E[\hat{\delta}^\ell_{jt}(s_t, \lambda) - x_{jt}\beta | z_{jt} = z]$ are continuous in the continuous components of z and the projection of $\mathcal{Z}_0$ onto the space of $z_{c,jt}$ (the continuous components of z_{jt}) is zero distance to an open set. This is innocuous in most applications.

Finally, Assumption 7 also requires that the instruments shift x_{jt} and π_{jt} sufficiently. This requirement is a standard one for BLP instruments. Thus, all the considerations for finding instruments in BLP models still apply.

The following theorem shows the consistency of the estimator defined in (26). Note that only the identified subvector θ^s_0 can be estimated consistently.

Theorem 1. *Suppose that either Assumption 1 holds and $T^{-1}\sum_{t=1}^{T} J_t^2/\bar{J}_T^2$ is bounded, or Assumption 2 holds and $\sup_{t=1,\dots,T} J_t$ is bounded. Further suppose that Assumptions 3–7 hold. Then $\|\widehat{\theta}^s_T - \theta^s_0\| \to_p 0$.*

Remark. Note that for logit-based models (which satisfy Assumption 1), we do not need J_t to be bounded. We require that the J_t's are roughly even across t, which is formalized as the boundedness of $T^{-1}\sum_{t=1}^{T} J_t/\bar{J}_T^2$. For nonlogit-based models satisfying Assumption 2, we require $\sup_{t=1,\dots,T} J_t$ to be bounded because Assumption 2(c) requires $\tilde{s}_{jt} = s_{jt} + 1/n_t$, which is incompatible with Assumption 4(f) unless $\max_{t=1,\dots,T} J_t$ is bounded.

Remark. The proof of the theorem follows two steps. First, we show that at the true value θ_0, $\widehat{Q}_T(\theta) = o_p(1)$. Second, we show that for points in Θ such that θ^s is bounded away from θ_0^s, $\widehat{Q}_T(\theta)$ asymptotically dominate $\widehat{Q}_T^*(\theta)$ and the latter is bounded away from zero. The proof is given in Section C.1.

7. Inference

In this section, we discuss statistical inference based on our point estimator. We show that the estimator is asymptotically normal despite that the bounds are slack for some g's, which is a similar result to that in Kahn and Tamer (2009) for censored regression models.[23]

Since the consistency is derived only for the subvector θ^s of θ, the asymptotically normality also will be about the subvector. For ease of notation, we consider the particular case where $\theta^s = (\lambda', \beta^{s,\prime})'$ where β^s is a subvector of β. The parameters in β excluded from β^s are the coefficients of variables that are zero whenever $z_{jt} \in \mathcal{Z}_0$.

More assumptions are needed. For clarity, we divide the assumptions into two groups, the first being standard ones similar to those in Freyberger (2015) and the second being the special assumptions that are needed to account for the presence and the unknown identity of the fringe products. Let $B_c(\lambda_0) = \{\lambda \in \Lambda : \|\lambda - \lambda_0\| \le c\}$ and $B_c(\pi_t) = \{\tilde{\pi}_t \in (0,1)^{J_t} : \mathbf{1}'\tilde{\pi}_t < 1, \|\pi_t - \tilde{\pi}_t\|_f \le c\}$. Let $\mathcal{G}\backslash\mathcal{G}_0$ denote the relative complement of $\mathcal{G}_0$ in $\mathcal{G}$. Let $\partial m_{jt}(\lambda)$ denote $\left(\partial\delta_{jt}(\pi_t, \lambda)/\partial\lambda' \quad x_{jt}^{s,\prime}\right)'$, where x_{jt}^s is the subvector of x_{jt} that correspond to β^s. Let

$$\Gamma_T(g) = (T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} E\big[\partial m_{jt}(\lambda_0) g(z_{jt})\big]. \tag{39}$$

Assumption 8. (a) *θ_0^s is in the interior of $\Theta^s := \{\theta^s : \exists \theta^r \text{ s.t. } (\theta^{s,\prime}, \theta^{r,\prime})' \in \Theta\}$.*

(b) *The function $\delta_{jt}(\pi, \lambda)$ is twice-continuously differentiable on $\Delta_{J_t}^0 \times \Lambda$, for all j, t.*

(c) *For some $c > 0$ and $M < \infty$,*

$$\sup_{j,t} E\left[\sup_{\tilde{\pi}_t \in B_c(\pi_t)} \sup_{\lambda \in B_c(\lambda_0)} \left\| \frac{\partial \delta_{jt}(\tilde{\pi}_t, \lambda)}{\partial \lambda} \right\|\right] \le M,$$

[23]However, it is worth noting some subtle differences between the identification and inference arguments in this paper and those in Kahn and Tamer. In Kahn and Tamer, the upper and lower bounds collapse in the *finite sample* for covariate values that indicate no-censoring, while in this paper, the upper and lower bounds collapse only asymptotically. Kahn and Tamer consider a fixed data-generating-process asymptotic framework, while the nature of our problem calls for a triangular array asymptotic framework. These are part of the reason that our conditions look more complicated than Kahn and Tamer's.

$$\sup_{j,t} E\left[\sup_{\lambda \in B_c(\lambda_0)} \left\| \frac{\partial^2 \delta_{jt}(\pi_t, \lambda)}{\partial \lambda \partial \lambda'} \right\|\right] \leq M,$$

and $\sup_{j,t} E[\|x^s_{jt}\|^2 | z_{jt} \in \mathcal{Z}_0] \leq M$.

(d) $\lim_{T\to\infty} \sum_{g\in\mathcal{G}_0} \mu(g)\Gamma_T(g)\Gamma_T(g)' = \Upsilon$ *for a matrix* Υ *of full rank, and*

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{g,g^*\in\mathcal{G}_0} \operatorname{Cov}\left(\bar{J}_T^{-1} \sum_{j=1}^{J_t} \xi_{jt} g(z_{jt}), \bar{J}_T^{-1} \sum_{j=1}^{J_t} \xi_{jt} g^*(z_{jt})\right) \Gamma_T(g)\Gamma_T(g)'\mu(g)\mu(g^*) = V.$$

(e) $\lim_{T\to\infty} T^{-1}\overline{n}_T^{1/2} = 0$.

Remark. Parts (a)–(b) are standard regularity conditions for extreme estimators. Part (c) imposes a uniform bound on the derivatives of $\delta_{jt}(\cdot, \lambda)$ with respect to λ. This bound condition is trivially satisfied for multinomial logit models and the binary probit models because $\delta_{jt}(\cdot, \lambda)$ does not depend on λ. For the nested logit model, $|\partial \delta_{jt}(\tilde{\pi}_t, \lambda)/\partial \lambda| = |\log(\tilde{\pi}_{gt}/\tilde{\pi}_{0t})| \leq 2|\log(\underline{\varepsilon}_0 - c)|$ as long as $\pi_{gt}, \pi_{0t} > \underline{\varepsilon}_{0t}$, and $c < \underline{\varepsilon}_0$. And $\partial^2 \delta_{jt}(\pi_t, \lambda)/\partial\lambda\partial\lambda' = 0$. Thus, part (c) holds if the share of each nest is bounded from zero. For mixed logit models, one can verify part (c) following similar arguments as those for Lemma S7 in Supplemental Appendix D, under the additional assumptions that the covariates with random coefficients are bounded. Part (d) of the assumption is needed, because we allow the data generating process to drift as $T \to \infty$. It regulates the limit of the drift in our asymptotic thought experiment. The only restriction it imposes on the data itself is that the Jacobian of the moment conditions has full-rank, which is standard for moment-based estimation and rules out perfect multicolinearity in x^s_{jt}.

Assumption 9. (a) *There exists a constant* $\eta > 0$ *such that for all sufficiently small* $c > 0$ *and all* T, *we have*

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:(T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} E[(\log(s_{jt}+\underline{\iota}_u/n_t)-\log(\pi_{jt}))g(z_{jt})]\leq c} \mu(g) < c^\eta,$$

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:(T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} E[(\log(s_{jt}+\bar{\iota}_\ell/n_t)-\log(\pi_{jt}))g(z_{jt})]\geq -c} \mu(g) < c^\eta,$$

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:(T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} E[g(z_{jt})(n_t s_{jt}+\iota_u)^{-1}]\leq c} \mu(g) < c^\eta.$$

(b) *Case* 1: *When Assumption* 1 *holds, assume that*

$$\sup_{j,t=1,\dots,T} E\left[\left\| \frac{\partial \check{\delta}_{jt}(\pi_t, \lambda_0)}{\partial \pi} \right\|^2\right] = O\big(J_T^{\max}\big) \quad \textit{and}$$

$$\sup_{j,t=1,\dots,T} \sup_{\pi\in B_c(\pi_t)} \left\| \frac{\partial^2 \check{\delta}_{jt}(\pi, \lambda_0)}{\partial \pi \partial \pi'} \right\| = O_p\big(J_T^{\max}\big)$$

for some $c > 0$.

Case 2: *When Assumption 2 holds, assume that*

$$\sup_{j,t=1,\dots,T} E\left[\left\|\frac{\partial \delta_{jt}(\pi_t,\lambda_0)}{\partial \pi}\right\|^2 1(z_{jt}\in\mathcal{Z}_0)\right]=O(1), \quad \textit{and}$$

$$\sup_{j,t=1,\dots,T}\ \sup_{\pi:\|\pi-\pi_t\|\le c}\left\|\frac{\partial^2\delta_{jt}(\pi,\lambda_0)}{\partial\pi\partial\pi'}1(z_{jt}\in\mathcal{Z}_0)\right\|=O_p(1)$$

for some $c>0$.

Remark. Part (a) of Assumption 9 is needed to show that the moments inequalities are slack enough for the fringe products to not interfere with the convergence rate and the asymptotic distribution of the bound estimator. It is satisfied for the $\mu(\cdot)$ and $\mathcal{G}$ that we propose if the exogenous variables that signal dominant products are discrete and π_{jt}'s for the fringe products converge to zero at the rate n_t^{-1} so that $E[\log(s_{jt}+\underline{\iota}_u/n_t)-\log\pi_{jt}|z_{jt}]$ and $E[\log(s_{jt}+\bar{\iota}_\ell/n_t)-\log\pi_{jt}|z_{jt}]$ are bounded away from zero. Part (b) strengthens the requirements of Assumptions 1 and 2 to ensure convergence rate of our estimator. The case 1 part of Assumption 9(b) implies the case 2 part of this assumption, thus is stronger. The weaker assumption is sufficient for case 2 because of the additional conditions in Assumption 2. Case 1 of Assumption 9(b) may be verified in a similar fashion as Assumption 1(a).[24]

Theorem 2. *Suppose that either Assumption 1 holds and* $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$ *is bounded, or Assumption 2 holds and* $\sup_{t=1,\dots,T} J_t$ *is bounded. Further suppose that Assumptions 3–9 hold. Then we have*

$$\sqrt{T}\big(\widehat{\theta}_T^s-\theta_0^s\big)\to_d N\big(0,\Upsilon^{-1}V\Upsilon^{-1}\big).$$

Remark 1. Note that Υ and V depend on $\mathcal{G}_0$, which in turn depends on the unknown set $\mathcal{Z}_0$. Thus, estimating the asymptotic variance covariance matrix can be difficult. Instead, following Kahn and Tamer (2009), we recommend using nonparametric bootstrap to obtain standard errors and confidence intervals. We follow this recommendation in the application in Section 9. We also evaluate the performance of bootstrap standard errors and bootstrap-based confidence intervals in our Monte Carlo experiments in Section 8.

Remark 2. The asymptotic variance formula also makes it clear that the choice of instrumental function set $\mathcal{G}$ affects estimation accuracy. Potentially, one could choose $\mathcal{G}$ to minimize the asymptotic variance, however, this does not seem to resemble the existing efficiency theory for conditional moment equalities, for example, Chamberlain (1987), Newey (1990), and Ai and Chen (2003), mainly due to the structure that $\mathcal{G}$ needs to take to preserve the information in the conditional moment inequalities. We thus leave this for future research.

[24] For multinomial logit and nested logit models, part (b) is not needed. The proofs of Theorem 2 go through with slight adjustment using the special structure of the inverse demand of such models, without using part (b). As a result, for such models the rate at which J_t increases with n_t does not need to be restricted.

8. Monte Carlo simulations

In this section, we present three sets of Monte Carlo experiments with random coefficient logit models. The first experiment investigates the performance of our approach with moderate fractions of zero shares, which should cover most of the empirical scenarios. In the second experiment, we test our estimator with a data generating process that produces extremely large fractions of zeros; the purpose is to further illustrate the key idea of our estimator in exploiting the long tail pattern that is naturally present in the data. In the third experiment, we use actual data from our application as the base for DGP; the purpose is to examine the performance of our estimator in a realistic setting and provide some practical guidelines regarding the choice of instruments functions.

The first two experiments use the a random coefficient logit model, where the utility of consumer i for product j in market t is

$$u_{ijt} = \alpha_0 + x_{jt}\beta_0 + \lambda_0 x_{jt} v_i + \xi_{jt} + \epsilon_{ijt}, \tag{40}$$

where $v_i \sim N(0, 1)$, λ_0 is the standard deviation of the random coefficients on x_{jt}, ϵ_{ijt}'s are i.i.d. across i, j, and t following Type I extreme value distribution. The parameters of interest are β_0 and λ_0, while α_0 is a nuisance parameter. In both experiments, we set $\lambda_0 = 0.5$, $\beta_0 = 1$, and vary α_0 for different designs. We simulate T markets, each with J products. For the third experiment, we will describe the DGP in Section 8.3.[25]

8.1 *Moderately many zeroes*

In the first experiment, the observed and unobserved characteristics are generated as $x_{jt} = \frac{j}{10} + N(0, 1)$ and $\xi_{jt} \sim N(0, 0.1^2)$ for each product j in market t. Thus, one feature of the design is that the x_{jt} has some persistence across markets—products with larger index tend to have higher value of x (which respects the nature of the variation in the scanner data shown in Section 2). Finally, the vector of empirical shares in market t, $(s_{0t}, s_{1t}, \dots, s_{Jt})$, is generated from Multinomial$(n, [\pi_{0t}, \pi_{1t}, \dots, \pi_{Jt}]')/n$, where n is the number of consumers in each market.[26]

With the simulated data set $\{(s_{jt}, x_{jt}) : j = 1, \dots, J\}_{t=1}^{T}$, we compute our bound estimator,[27] the standard BLP estimator using empirical share s_t in place of π_t and dis-

[25]In all of the three experiments, we checked the realized $\min_{j,t} n_t\pi_{jt}$'s in the generated data and they are all well above our choice of $\iota_\ell = 2^{-52} = 2.2204e-16$. Thus, there is a $\underline{\varepsilon}_1$ that is well above $\iota_\ell = 2^{-52}$ and satisfies $\min_{j,t} n_t\pi_{jt} \geq \underline{\varepsilon}_1$. This and Table 2 imply that our choice of ι_ℓ satisfies $0 < \iota_\ell \leq \overline{\iota}_\ell$, which is the key for the validity of the lower bound part of our moment inequality construction.

[26]The π_t has no closed-form solution in the random coefficient model, and thus we compute them via simulation, that is,

$$\pi_{jt} = \frac{1}{s}\sum_{i=1}^{s} \frac{\exp(\alpha_0 + x_{jt}\beta_0 + \lambda_0 x_{jt} v_i + \xi_{jt})}{1 + \sum_{k=1}^{J}\exp(\alpha_0 + x_{kt}\beta_0 + \lambda_0 x_{kt} v_i + \xi_{kt})},$$

where $s = 1000$ is the number of consumer type draws (v_i).

[27]We use $\tilde{s}_{jt} = s_{jt} + 1/(n_t J_t)$ for $j = 1, \dots, J_t$ when implementing the bound estimator for all the simulations in this section.

carding observations with $s_{jt} = 0$, the standard BLP estimator using Laplace shares $s_t^L = (n_t s_t + 1)/(n_t + J_t + 1)$ in place of π_t.

All the estimators require simulating the market shares and solving demand systems for each trial of λ in optimizing the objective function for estimation. We use the same set of random draws of v_i in estimation as in the data generating process to eliminate simulation error as simulation error is not the focus of this paper. BLP contraction mapping method is employed to numerically solve the demand systems for all three estimators.

We simulate 1000 data sets $\{(s_t^r, x_t^r) : t = 1, \dots, T\}_{r=1}^{1000}$ and implement all the estimators mentioned above on each for a repeated simulation study. For the instrumental functions, we use the countable hypercubes defined in (16), and set $\bar{r}_T = 50$. The choices of ι_ℓ and ι_u follow the recommendation in Section 4.2. For the BLP estimator, we use $(1, x_{jt}, x_{jt}^2 - 1, x_{jt}^3 - 3x_{jt})$ (the first three Hermite polynomials) as instruments to construct the GMM objective function. Alternative transformations of x_{jt} as instruments yield effectively the same results.

The bias and standard deviation of the estimators are presented in Table 3. As we can see from the table, the standard BLP estimator with using empirical share s_t (labeled as "ES") shows large bias for both β and λ. Replacing the empirical share s_t with the Laplace share s_t^L (and thus not discarding the observations with $s_{jt} = 0$), labeled as "LS," increases the bias for β although reducing the bias for λ. Our bound estimator (labeled as "Bound") is the least biased, and its bias is very small for both parameters, especially when the sample size (T) is large.

Next, we examine the performance of our proposed bootstrap procedure and the results are reported in Table 4. We can see that bootstrap standard errors are on average slightly larger than the standard deviation of the estimators, especially for the cases with large fraction of zeros and small sample size. Also, we compute two versions of bootstrap confidence intervals and find that the "Normal CI," based on normal quantile and bootstrap standard errors, outperforms the standard nonparametric percentile bootstrap confidence interval and gets rather close to the nominal level (95%) of coverage probability, especially for the β, when the sample size gets large and the fraction of zeros is not too high.

8.2 *Extremely many zeroes*

Next, we pressure test our bound estimator by pushing the fraction of zeroes in empirical shares toward the extreme. We modify the DGP slightly to produce a very high fraction of zeros. Specifically, we generate x_{jt} from the following discrete distribution:

x	1	12	15
$\Pr(x_{jt} = x)$	0.99	0.005	0.005

and

$$\xi_{jt} \sim 1(x_{jt} = 1) \times N(0, 2^2) + 1(x_{jt} \neq 1) \times N(0, 0.1^2).$$

All the other aspects of the DGP is identical to the previous simulation.

Table 3. Monte Carlo results: estimation.

DGP	T	Ave. %of Zeros		ES λ	ES β	LS λ	LS β	Bound λ	Bound β
I	25	9.52%	Bias	0.3718	−0.1941	0.2900	−0.2167	0.0422	−0.0432
			SD	0.0337	0.0160	0.0221	0.0115	0.0477	0.0352
	50	9.48%	Bias	0.3712	−0.1939	0.2912	−0.2172	0.0172	−0.0216
			SD	0.0236	0.0118	0.0164	0.0082	0.0388	0.0284
	100	9.46%	Bias	0.3714	−0.1941	0.2900	−0.2169	0.0002	−0.0065
			SD	0.0169	0.0081	0.0112	0.0055	0.0311	0.0234
II	25	18.54%	Bias	0.6752	−0.6115	0.4023	−0.4675	0.0142	−0.0302
			SD	0.0845	0.0655	0.0315	0.0229	0.0531	0.0536
	50	18.54%	Bias	0.6649	−0.6040	0.3993	−0.4657	−0.0083	−0.0028
			SD	0.0580	0.0462	0.0223	0.0158	0.0410	0.0413
	100	18.50%	Bias	0.6624	−0.6021	0.3983	−0.4651	−0.0154	0.0073
			SD	0.0422	0.0333	0.0163	0.0114	0.0297	0.0297
III	25	41.13%	Bias	0.7302	−1.3220	0.3868	−0.9863	−0.0366	0.0278
			SD	0.2022	0.2890	0.0366	0.0460	0.0481	0.0721
	50	41.09%	Bias	0.7092	−1.2947	0.3830	−0.9819	−0.0331	0.0303
			SD	0.1373	0.1975	0.0262	0.0323	0.0374	0.0549
	100	41.09%	Bias	0.7070	−1.2935	0.3809	−0.9794	−0.0219	0.0176
			SD	0.0911	0.1325	0.0188	0.0232	0.0282	0.0391
IV	25	52.39%	Bias	0.4013	−1.1035	0.2907	−1.1412	−0.0499	0.0512
			SD	0.1346	0.2435	0.0304	0.0453	0.0530	0.0899
	50	52.35%	Bias	0.3942	−1.0937	0.2877	−1.1369	−0.0346	0.0330
			SD	0.0956	0.1740	0.0214	0.0313	0.0396	0.0635
	100	52.36%	Bias	0.3916	−1.0901	0.2862	−1.1349	−0.0215	0.0169
			SD	0.0687	0.1255	0.0154	0.0227	0.0311	0.0475

Note: 1. $J = 50$, $n = 10{,}000$, $\beta_0 = 1$, $\lambda_0 = 0.5$, number of repetitions = 1000. 2. "ES": Empirical Shares; "LS": Laplace Shares. 3. DGP: I, II, III, and IV correspond to $\alpha_0 = -9, -10, -12$, and -13, respectively.

The fractions of zeroes are made very high: 82%–96% by choosing the α_0 parameter. With such high fractions of zeroes, the vast majority of observations are uninformative. Thus, we need larger sample size for any estimator to perform well. We consider $T = 100, 200, 400$. For simplicity of presentation and to reduce computational burden, here we fix λ at its true value, and only investigate the behavior of the estimators for β.

The results are reported in Table 5, and they are very encouraging for the bound approach. The ES estimator is severely biased toward 0, so is the LS estimator. The bound estimator is remarkably accurate in these extreme cases. The performance highlights the key idea behind our estimator: utilizing the information from the dominant products with inherently thick demand while controlling the impact of fringe products with small/zero sales on estimation.

8.3 *Monte Carlo simulations with tuna data*

In this subsection, we conduct Monte Carlos simulations based on the canned tuna data set that will be used later in our application. The main purposes are two-fold: (1) we want

Table 4. Monte Carlo results: bootstrap.

DGP	T	Ave. %of Zeros	Actual SD λ	Actual SD β	BS SE λ	BS SE β	CP: BS CI λ	CP: BS CI β	CP: Normal CI λ	CP: Normal CI β
	25	9.52%	0.0477	0.0352	0.0473	0.0353	0.8390	0.8250	0.8630	0.7790
I	50	9.48%	0.0388	0.0284	0.0400	0.0300	0.8556	0.8675	0.9444	0.9160
	100	9.46%	0.0311	0.0234	0.0324	0.0244	0.8408	0.8458	0.9570	0.9530
	25	18.54%	0.0531	0.0536	0.0563	0.0585	0.8390	0.8630	0.9640	0.9510
II	50	18.54%	0.0410	0.0413	0.0423	0.0433	0.7980	0.8340	0.9490	0.9690
	100	18.50%	0.0297	0.0297	0.0309	0.0311	0.8380	0.8750	0.9270	0.9560
	25	41.13%	0.0481	0.0721	0.0537	0.0840	0.7700	0.8310	0.9040	0.9680
III	50	41.09%	0.0374	0.0549	0.0388	0.0581	0.8360	0.8760	0.8690	0.9360
	100	41.09%	0.0282	0.0391	0.0290	0.0417	0.8740	0.9250	0.9000	0.9450
	25	52.39%	0.0530	0.0899	0.0549	0.0971	0.7880	0.8550	0.8710	0.9430
IV	50	52.35%	0.0396	0.0635	0.0420	0.0707	0.8490	0.9120	0.8870	0.9530
	100	52.36%	0.0311	0.0475	0.0312	0.0498	0.8450	0.8980	0.9040	0.9440

Note: 1. All the settings are identical to Table 1. Bootstrap draws are taken at market level. Bootstrap sample size is 500. 2. “BS SE” refers to average bootstrap standard error. 3. “CP: BS CI” refers to the coverage probability of the 95% nonparametric bootstrap CI. 4. “CP: Normal CI” refers to the coverage probability of the 95% normal CI with bootstrap s.e.

to examine performance of the bound estimator in a setting that is similar to the application; (2) we would like to understand better how the choices of instruments affect the performance of the bound estimator, especially in real empirical settings where product dummies are typically included.

Table 5. Monte Carlo results: very large fraction of zeros.

DGP	T	Avg. % of Zeros		ES	LS	Bound
I	100	84.73%	Bias	−0.2698	−0.2643	−0.0014
			SD	0.0060	0.0058	0.0123
	200	84.68%	Bias	−0.2695	−0.2640	−0.0011
			SD	0.0042	0.0040	0.0094
	400	84.71%	Bias	−0.2692	−0.2639	−0.0005
			SD	0.0030	0.0030	0.0066
II	100	91.45%	Bias	−0.3328	−0.3319	−0.0016
			SD	0.0066	0.0061	0.0126
	200	91.43%	Bias	−0.3324	−0.3314	−0.0014
			SD	0.0049	0.0044	0.0091
	400	91.43%	Bias	−0.3320	−0.3313	−0.0007
			SD	0.0034	0.0032	0.0067
III	100	95.37%	Bias	−0.3992	−0.4028	−0.0014
			SD	0.0079	0.0070	0.0126
	200	95.35%	Bias	−0.3991	−0.4025	−0.0014
			SD	0.0056	0.0049	0.0093
	400	95.36%	Bias	−0.3986	−0.4023	−0.0010
			SD	0.0040	0.0035	0.0065

Note: 1. $T = 100$, $J = 50$, $n = 10{,}000$, $\beta_0 = 1$, $\lambda_0 = 0.5$. Number of repetitions = 1000. 2. We fix $\lambda = \lambda_0$ (at the true value) without estimating it. 3. DGP: I, II, III correspond to $\alpha_0 = -13, -14, -15$.

To generate data, we use tuna data in one week (the week of March 30, 1995) across all the stores (there are 80 stores) as a template (a store week as a "market" and a UPC as a "product") and consider a random coefficient logit specification that extends (40). In particular, we let the price coefficient be random, that is,

$$u_{ijt} = a_0 + \beta_0 x_{jt} - v_i p_{jt} + \xi_{jt} + \epsilon_{ijt},$$

where v_i follows lognormal(μ_p, σ_p). The product-market specific demand shock ξ_{jt} has a simple heteroskedasticity structure

$$\xi_{jt} = 1\big(\beta_0 x_{jt} \geq \mathrm{med}(\beta_0 x_{jt})\big)\xi'_{jt} + 1\big(\beta_0 x_{jt} < \mathrm{med}(\beta_0 x_{jt})\big)\xi''_{jt},$$

where ξ'_{jt} (ξ''_{jt}) follows normal distribution $N(0, 0.5^2)$ ($N(0, 1.5^2)$) truncated at $\pm 3\sigma$. The truncation gives ξ_{jt} a finite support to ensure that Assumptions 3 and 5 hold easily. Price is generated as a linear combination of marginal cost (use the observed wholesale price from the data) and a markup term that is a function of demand shock ξ, that is,

$$p_{jt} = mc_{jt} + b_0 \exp(\xi_{jt}). \tag{41}$$

Note that the markup term introduces a simple endogeneity problem. The covariates x_{jt} include a continuous variable following $N(0, 1)$ (truncated at $\pm 3\sigma$) and UPC dummies from the data. The coefficient on the continuous variable is 1 and those on the UPC dummies are set to be the estimated ones (using bound estimator) in our application. Other specifications are similar to the previous DGP. And the number of consumers in each market for generating market shares is directly imported from the data.

We simulate 1000 data sets that have the same structure as the real data, with the endogenous variables, that is, price and market shares, varying across data sets. Then we implement several estimators of interests using the data sets. To simplify the estimation, we only estimate the two parameters of the random coefficient on price and fixing other parameters (UPC fixed effects) at their true values without estimating them.

The estimation results are summarized in Table 6. Note that we consider three values of a_0 that imply different fractions of zeros (labeled by "I," "II," and "III"). Also, besides the baseline $T = 80$ case with one week data (the week of March 30, 1995), we also try $T = 160$ using two weeks' data (the weeks of March 23 and March 30, 1995). As before, "ES" and "LS" refer to standard BLP estimator applied to empirical shares and Laplace shares, respectively. For the bound estimator, we experiment with four alternative sets of instrument functions. "Bound-$\mathcal{G}_1$" uses the instruments defined by (16), which includes indicators constructed from continuous variables (z_{jt}, mc_{jt}) with $\bar{r}_{80} = 10$ and $\bar{r}_{160} = 15$ and UPC dummies. "Bound-$\mathcal{G}_2$" is the same as "Bound-$\mathcal{G}_1$" except with larger $\bar{r}_T$: $\bar{r}_{80} = 20$ and $\bar{r}_{160} = 30$. "Bound-$\mathcal{G}_3$" ("Bound-$\mathcal{G}_4$") expands the set of instruments of "Bound-$\mathcal{G}_1$" ("Bound-$\mathcal{G}_2$") by including the interactions between indicators constructed from continuous variables (denoted by $\mathcal{C}$ in (16)) and UPC dummies.

From the results, we can see that

- In almost all the cases, as before, the bound estimators have much smaller biases than the ES and LS estimators do (although with slightly increased standard deviations), especially for the standard deviation of the random coefficient σ_p.

TABLE 6. Monte Carlo results: simulation using tuna data.

DGP	T	Avg. % of Zeros		ES	LS	Bound-$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_3$	$\mathcal{G}_4$
						Panel I: μ_p			
I	80	9.11%	Bias	−0.0001	−0.0605	−0.0009	−0.0007	−0.0004	−0.0014
			SD	0.0228	0.0251	0.0326	0.0239	0.0299	0.0233
	160	9.09%	Bias	−0.0081	−0.0678	−0.0026	−0.0042	−0.0019	−0.0037
			SD	0.0158	0.0185	0.0296	0.0175	0.0276	0.0173
II	80	14.29%	Bias	−0.0166	−0.1403	−0.0098	−0.0104	−0.0096	−0.0099
			SD	0.0258	0.0311	0.0582	0.0303	0.0498	0.0305
	160	14.25%	Bias	−0.0246	−0.1472	−0.0005	−0.0129	−0.0037	−0.0126
			SD	0.0186	0.0228	0.0721	0.0221	0.0530	0.0220
III	80	17.70%	Bias	−0.0286	−0.2205	−0.0097	−0.0185	−0.0123	−0.0180
			SD	0.0269	0.0318	0.0767	0.0344	0.0635	0.0344
	160	17.66%	Bias	−0.0373	−0.2267	0.0203	−0.0185	−0.0013	−0.0183
			SD	0.0192	0.0234	0.1178	0.0253	0.0706	0.0252
						Panel II: σ_p			
I	80	9.11%	Bias	0.2691	0.4111	0.0555	0.0840	0.0630	0.0840
			SD	0.0664	0.0848	0.1382	0.0846	0.1176	0.0804
	160	9.09%	Bias	0.2457	0.3931	0.0065	0.0402	0.0208	0.0414
			SD	0.0484	0.0647	0.1570	0.0662	0.1338	0.0657
II	80	14.29%	Bias	0.3103	0.5125	0.0163	0.0656	0.0367	0.0664
			SD	0.0674	0.1030	0.2469	0.0978	0.2119	0.0990
	160	14.25%	Bias	0.2924	0.5022	0.0007	0.0290	0.0324	0.0304
			SD	0.0511	0.0778	0.3153	0.0747	0.2633	0.0747
III	80	17.70%	Bias	0.3410	0.5745	0.0340	0.0578	0.0458	0.0590
			SD	0.0710	0.1164	0.3273	0.1163	0.2873	0.1179
	160	17.66%	Bias	0.3221	0.5678	0.1010	0.0260	0.1158	0.0275
			SD	0.0517	0.0871	0.3819	0.0911	0.2915	0.0912

Note: 1. DGP: I, II, and III correspond to $a_0 = 0.4, 0.8, 1$, respectively. Number of markets T: 80 and 160 correspond to one week (03/30/1995 to 04/05/1995) and two weeks (03/23/1995 to 04/05/1995) of the tuna data for all the stores, respectively. 2. "E": Empirical Shares; "LS": Laplace Shares; "Bound-$\mathcal{G}_1$": $\bar{r}_{80} = 10$, $\bar{r}_{160} = 15$; "Bound-$\mathcal{G}_2$": $\bar{r}_{80} = 20$, $\bar{r}_{160} = 30$; "Bound-$\mathcal{G}_3$": $\bar{r}_{80} = 10$, $\bar{r}_{160} = 15$, instruments in $\mathcal{C}$ interact with product dummies; "Bound-$\mathcal{G}_4$": $\bar{r}_{80} = 20$, $\bar{r}_{160} = 30$, instruments in $\mathcal{C}$ interact with product dummies. 3. True value: $\mu_p = 0$, $\sigma_p = 0.5$. Coefficients on product dummies are fixed at their true values without being estimated for ease of computation. Number of repetitions = 1000.

- By comparing Bound-$\mathcal{G}_1$ and Bound-$\mathcal{G}_2$, we can see that increasing the tuning parameter $\bar{r}_T$ reduces standard deviation substantially but increase biases slightly.
- Including interactions between $\mathcal{C}$ and UPC dummies reduces standard deviations of the estimators substantially. Hence, it seems preferable to have a sufficiently large $\bar{r}_T$ and include the interactions, and these findings guide the construction of $\mathcal{G}$ in our empirical application.

9. Empirical application

In this section, we apply our estimator on the DFF scanner data previewed in Section 2. [28] In particular, we focus on the canned tuna category, as previously studied by Chevalier, Kashyap, and Rossi (2003) (CKR for short) and Nevo and Hatzitaskos (2006) (NH for short). CKR observed using the same data discussed in Section 2 that the share-weighted average price (i.e., the price index) of tuna fell by 15% during Lent—a high demand period for this product. They attributed the outcome to loss-leading behavior on the part of retailers. NH on the other hand suggest that this pricing pattern in the tuna data could instead be explained by increased price sensitivity of consumers (consistent with an increase in search), which causes a reallocation of market shares toward less expensive products in the Lent period, and hence a fall in the observed share weighted price index. They test this hypothesis directly in the data by estimating demand parameters separately in the Lent and non-Lent periods, and find that demand becomes more elastic in the high demand (Lent) period.

Here, we revisit the groundwork laid by NH to examine the difference in price elasticity between Lent and non-Lent periods. The main difference in our analysis is that we use data on all products in the analysis, while NH restrict the sample to include only the top 30 UPCs, and thus automatically drop products with small/zero sales. There are two main questions we seek to address: (a) Does the selection of UPCs with only positive shares significantly bias the estimates of price elasticity and (b) Does the difference in price elasticities between the Lent and non-Lent period persist after properly controlling for zeroes.

To make the comparison clear, we use largely the same specification of the model used in NH. In particular, we consider a logit specification

$$u_{ijt} = \alpha p_{jt} + \beta x_{jt} + \xi_{jt} + \epsilon_{ijt},$$

where the control variables x_{jt} consist of UPC fixed effects and a time trend.[29] The week-to- week variation in the product-/market-level unobserved demand shock ξ_{jt} largely captures the short-term promotional efforts, for example, in-store advertising and shelving choices, because the UPC fixed effects control the intrinsic product quality, which is likely stable over a short time horizon. Since stores are likely to advertise or shelf the product in a more prominent way during weeks when the product is on a price sale, we expect a negative correlation between price and the unobservable. We construct instruments for price by inverting DFF's data on gross margin to calculate the chain's wholesale costs, which is the standard price instrument in the literature that has studied the DFF data.[30]

[28] The sample period predates the price fixing conduct by the tuna cartel starting around 2011; see Miller, Remer, and Weinberg (2020) for details.

[29] Empirical market shares are constructed using quantity sales and the number of people who visited the store that week (the customer count) as the relevant market size.

[30] The gross margin is defined as (retail price–wholesale cost)/retail price, so we get wholesale cost using retail price × (1 − gross margin). The instrument is defensible in the store disaggregated context we consider here because it has been shown that price sales in retail price primarily reflect a reduction in retailer mar-

Table 7. Demand estimation results.

	BLP	Bound
Price coefficient	−0.39	−1.03
S.E.	(0.005)	(0.319)
Avg. own price elasticity	−0.57	−1.51
Fraction of inelastic products	90.04%	28.20%
No. of obs.	862,683	959,331

Note: The S.E. for the bound approach is the bootstrap standard error (using 1000 bootstrap replications).

We implement our bound estimator defined by (26) to obtain the point estimate of (α, β) in the model.[31] The standard errors are obtained using nonparametric bootstrap.[32] The estimation results are presented in Tables 7 and 8.[33] Table 7 shows that standard BLP logit estimator that inverts empirical shares to recover mean utilities (and hence drops zeroes) has a significant selection bias toward zero. The UPC level elasticities for the logit model are small in economic magnitude, with the average elasticity in the data being −0.572. Furthermore, over 90% of products have inelastic demand. Using our bounds approach instead to control for zeroes has a major effect on the estimated elasticities. Average demand elasticity for UPCs becomes −1.51 and less than 30% of observations have inelastic demand. This change in the direction of elasticities is consistent with the attenuation bias effects of dropping products with small/zero market shares.

Our second result is that demand becomes more elastic in the high demand period, as shown in Table 8. This is consistent with Nevo and Hatzitaskos (2006)'s findings that are based on the standard logit estimator with zeroes being dropped. However, the Lent effect is bigger according to our bounds estimator that controls for the zeroes. In other words, correcting the selection bias, our bound estimator brings the price coefficient and elasticity higher and the correction effect is higher for the Lent period than for the non-Lent period. Since the fractions of zeroes are remarkably close between Lent and

gins rather than a reduction in wholesale costs (see, e.g., Chevalier, Kashyap, and Rossi (2003) and Hosken and Reiffen (2004)); thus, sales (and hence promotions) are not being driven by the manufacturer through temporary reduction in wholesale costs. However, this instrument may be invalid if manufacturers respond to demand shocks and adjust wholesale prices accordingly. We acknowledge the potential deficiency of using this instrument but searching for a better alternative is beyond the scope of the current paper.

[31]The choice of $\mathcal{G}$ is guided by the simulation results in Section 8.3: we set $\bar{r} = 45$ when constructing instrument functions from the wholesale cost (continuous variable) and include interactions between them and the UPC dummies.

[32]The procedure contains the following steps: (1) draw with replacement a bootstrap sample of *markets*, denoted as $\{t_1, \dots, t_T\}$; (2) compute the bound estimator $\widehat{\theta}_T^{BD*}$ using the bootstrap sample; (3) repeat (1)–(2) for B_T times and obtain B_T independent (conditional on the original sample) copies of $\widehat{\theta}_T^{BD*}$; (4) obtain the sample standard deviation of the B_T copies of $\widehat{\theta}_T^{BD*}$ and this is the bootstrap standard error.

[33]In principle, we can estimate our model separately for each store, letting preferences change freely over stores depending on local preferences. These results are available upon request. Here, we present for the results of demand pooling together all stores together as was done by Nevo and Hatzitaskos (2006). The store level regressions results are very similar to the pooled store regression and the latter is a more concise summary of demand behavior that we present here.

Table 8. Demand in Lent versus non-Lent.

	BLP		Bound	
	Lent	Non-Lent	Lent	Non-Lent
Price coefficient	−0.518	−0.371	−1.23	−0.75
S.E.	(0.018)	(0.005)	(0.221)	(0.231)
Avg. own price elasticity	−0.757	−0.544	−1.80	−1.10
Fraction of inelastic products	84.02%	92.84%	16.79%	43.94%
No. of obs.	70,496	792,187	78,838	880,493

Note: The s.e. for the bound approach is the bootstrap standard error (using 1000 bootstrap replications).

non-Lent periods, we suspect that the difference in the correction effect is due to a difference in the distribution of the unobservable ξ.

To further investigate this, we first replicate the reduced form finding of Nevo and Hatzitaskos (2006) that suggested a change in price sensitivity in the Lent period. This is reported in Table 9, which shows that although the price index of tuna during Lent appears to be approximately 15% less expensive than other weeks (as previously underscored by CKR), the average price of tuna is virtually unchanged between the Lent versus non-Lent period. Hence, it is a reallocation of demand toward less expensive products during Lent that drives the change in the price index.

We take this decomposition one step further than NH, and examine the price index separately for products "on sale" and "regularly priced" during these periods.[34] As can be seen in Table 10, it is the sales price index that is the key driver of the aggregate price index being cheaper during Lent. However, the average price of an "on-sale" product is not cheaper in the Lent period. This shows that it is a reallocation toward more steeply discounted "on-sale" product during Lent that is driving change in the aggregate price index. In contrast, we do not see an analogous reallocation for "regularly priced" products.

This suggests a tighter coordination of promotional effort and discounting in the high- demand period. In effect, more steeply discounted products are receiving larger promotional effort on the part of the retailer during the high demand, which is similar in spirit to the loss-leader hypothesis originally advanced for this data by CKR. Since promotional effort in the model is largely captured through the unobservable ξ, this change in behavior of the unobservable would account for the selection effect due to

Table 9. Regression of price index in Lent.

	P: Price Index	$\bar{P}$: Average Price
Lent	−0.150	−0.009
s.e.	(0.0005)	(0.0003)

[34]We flag an observation in the data as being on sale if that particular UPC in that particular store in that particular week has at least a 5% reduction from highest price of previous three weeks.

Table 10. Regression of sales price index in Lent.

	P: Price Index		$\bar{P}$: Average Price	
	Sale	Regular	Sale	Regular
Lent	−0.199	0.035	0.010	0.001
s.e.	(0.0017)	(0.0003)	(0.0016)	(0.0003)

dropping zeroes changing across the two periods: during the Lent period, the variance of promotional effort is larger so the selection bias is worse. Hence, our results suggest that both demand and supply side effects contribute to the falling price during the high-demand period, which complements and reconciles the findings of NH and CKR.

10. Conclusion

We have shown that differentiated product demand models have enough content to construct a system of moment inequalities that can be used to consistently estimate demand parameters despite a possibly large presence of observations with zero market shares in the data. We construct a GMM-type estimator based on these moment inequalities that is consistent and asymptotically normal under assumptions that are a reasonable approximation to the DGP in various environments with product differentiation. Our application to scanner data reveals that taking the market zeroes in the data into account has economically important implications for price elasticities.

References

Ai, Chunrong and Xiaohong Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71, 1795–1843. [406]

Anderson, Chris (2006), *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion. [384]

Andrews, Donald W. K. and Xiaoxia Shi (2013), "Inference based on conditional moment inequality models." *Econometrica*, 81. [391, 392, 398, 401, 402]

Armstrong, Timothy B. (2016), "Large market asymptotics for differentiated product demand estimators with economic models of supply." *Econometrica*, 84, 1961–1980. [387]

Berry, Steve, Oliver B. Linton, and Ariel Pakes (2004), "Limit theorems for estimating the parameters of differentiated product demand systems." *Review of Economic Studies*, 71 (3), 613–654. [387, 388, 401]

Berry, Steven, Amit Gandhi, and Philip Haile (2013), "Connected substitutes and invertibility of demand." *Econometrica*, 81 (5), 2087–2111. [393]

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*, 841–890. [382, 387]

Berry, Steven and Ariel Pakes (2007), "The pure characteristic demand model." *International Economic Review*, 48 (4), 1193–1225. [393]

Berry, Steven T. (1994), "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262. [386, 387]

Chamberlain, Gary (1987), "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34, 305–334. [406]

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003), "Why don't prices rise during periods of peak demand? Evidence from scanner data." *American Economic Review*, 93 (1), 15–37. [383, 413, 414]

Cox, Gregory and Xiaoxia Shi (2019), "Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models." Unpublished manuscript, Department of Economics, University of Wisconsin at Madison. [391, 401]

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su (2012), "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation." *Econometrica*, 80 (5), 2231–2267. [383]

Dube, Jean-Pierre, Ali Hortacsu, and Joonhwi Joo (2020), "Random-coefficients logit demand estimation with zero-valued market shares." Becker Friedman Institute working paper No. 2020-13. [382]

Freyberger, Joachim (2015), "Asymptotic theory for differentiated products demand models with many markets." *Journal of Econometrics*, 185 (1), 162–181. [387, 388, 401, 402, 404]

Gabaix, Xavier (1999), "Zipf's law and the growth of cities." *The American Economic Review, Papers and Proceedings*, 89, 129–132. [384]

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi (2013), "Estimating demand for differentiated products with error in market shares." CeMMAP working paper. [401]

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi (2023), "Supplement to 'Estimating demand for differentiated products with zeroes in market share data'." *Quantitative Economics Supplemental Material*, 14, https://doi.org/10.3982/QE1593. [384]

Gowrisankaran, Gautam and Thomas J. Holmes (2004), "Mergers and the evolution of industry concentration: Result from the dominant firm model." 35, 561–582. [398]

Hosken, Daniel and David Reiffen (2004), "Patterns of retail price variation." *RAND Journal of Economics*, 128–146. [414]

Kahn, Shakeeb and Elie Tamer (2009), "Inference on randomly censored regression models using conditional moment inequalities." *Journal of Econometrics*, 152, 104–119. [383, 391, 404, 406]

Lima, Lucas (2021), "Demand estimation with zeros: Moving costs in the US." Harvard University working paper. [382]

Markham, Jesse W. (1951), "The nature and significance of price leadership." *The American Economic Review*, 41, 891–905. [398]

Miller, Nathan H., Marc Remer, and Matthew C. Weinberg (2020), "The canned tuna cartel." Discussion paper. [413]

Nevo, Aviv and Konstantinos Hatzitaskos (2006), "Why does the average price paid fall during high demand periods?" Discussion paper, CSIO working paper. [383, 413, 414, 415]

Newey, Whitney K. (1990), "Efficient instrumental variables estimation of nonlinear models." *Econometrica*, 58, 809–837. [406]

Nurski, Laura and Frank Verboven (2016), "Exclusive dealing as a barrier to entry? Evidence from automobiles." *The Review of Economic Studies*, 83 (3), 1156. [382]

Quan, Thomas W. and Kevin R. Williams (2018), "Product variety, across-market demand heterogeneity, and the value of online retail." *The RAND Journal of Economics*, 49 (4), 877–913. [382]

Shimomura, Ken-Ichi and Jacques-Francois Thisse (2012), "Competition among the big and the small." *The RAND Journal of Economics*, 43, 329–347. [387, 398]

Tirole, Jean (1988), *The Theory of Industrial Organization*. The MIT Press. [398]

Co-editor Andres Santos handled this manuscript.

Manuscript received 6 April, 2020; final version accepted 8 August, 2022; available online 9 September, 2022.