Demand Estimation with Scanner Data: Revisiting the Loss-Leader Hypothesis

Amit Gandhi, Zhentong Lu, Xiaoxia Shi
University of Wisconsin-Madison
February 28, 2014

Abstract

Why are retail prices frequently discounted? The loss-leader model suggests that promotion is an important reason. In this paper we build and estimate an empirical model of demand facing a retailer to shed light on this explanation. There are two key features of our demand model that are novel. The first is that we model demand at the finest level of disaggregation in scanner data - the demand for an SKU at a single store in a given week. The second is that we allow for a demand unobservable to simultaneously shift with the price reduction, which will capture promotional effort that is unobserved to the econometrician. Estimating demand at this disaggregate level is challenging because of the large prevalence of “market zeroes” - products that do not sell in a given store at a given week. We construct optimal market share for demand estimation that shrink empirical market shares and thus solves the zeroes problem. We apply our approach to tuna prices and quantities in scanner data and show the apparent counter-cyclical behavior of prices is consistent with loss leader behavior by the retailer.

1 Introduction

Scanner data provides a unique window into the behavior of retail prices. A fundamental pattern in scanner data that has been documented by several studies is that products exhibit “sticky” regular prices and frequent discounts relative to the regular price (see e.g., Hosken and Reiffen [2004], Nakamura and Steinsson [2008]). We illustrate this canonical pattern below in Figure 1.1 for the top selling SKU (the bar code that identifies a unique product) of bathroom tissue at a single store in our data. The X-axis represents the weeks in the data and the Y-axis depicts both standardized prices and quantities sold for this SKU.
Why does demand spike when there is a temporary price reduction? The most obvious answer is that demand increases because price decreases. That is, we are observing a movement down a demand curve.

Yet further reflection makes the answer less obvious. A retailer does not passively discount a product but rather actively promotes products that are discounted. This bundling of discounts and promotions is is the basis for the loss-leader theory for retail sales - promoting the products on sale sends a credible signal to consumers about prices they will face, which in turn facilitates competition among stores (see e.g., Lal and Matutes [1994]).

Despite its intuitiveness and general acceptance among practitioners, the loss leader theory has been subject to very little empirical analysis in the economics literature[^1] If the hypothesis is correct then at least some of the demand spike we see associated with price reductions is due to the promotional effort/marketing. This is the key empirical question we aim to address in this paper: How much of the demand response to a store discounting a product is coming from the reduction in price (movement along a demand curve) versus promotion (a shift in the demand curve due to marketing). To address this question we build and estimate an empirical model of demand facing a retailer where the retailer both prices its products and exerts promotional effort in the marketing.

There are two key difficulties we face. The first difficulty that promotional effort poses for empirical work is that it is largely unobserved and often hard to quantify. Although scanner data will typically offer some indicators of promotion by the retailer, these measures are insufficient to capture the full extent promotional effort. For example the exact position of a product on a shelf is difficult to capture in a form that is amenable to empirical analysis yet is known to play a major role in purchasing behavior (see e.g., Chandon et al. [2009]). Thus we must account for demand

[^1]: Notable exceptions include Chevalier et al. [2003] and Ellison and Ellison [2009]. We will later discuss the results of former paper.
unobservables that partly reflect these unobserved effects of promotions. However we are now faced
with a classic “supply and demand” simultaneity problem in isolating the price effect on demand -
demand does not stay constant when price changes.

The second key difficulty is that we must model demand at the finest level of disaggregation
in scanner data - the demand for an SKU at a single store in a given week. Modeling demand at
this disaggregated level is essential for studying the variation in Figure 1.1 - any aggregation over
products, stores, or weeks will artificially eliminates the actual pattern of sales in the data. Yet
at this highly disaggregated level we encounter a serious empirical problem - many products in a
store during a given week simply do not sell. That is, the aggregate demand for many products
in a marketing period is zero. Such “market zeroes” arise in many economic environments when
markets and products are disaggregated.2

Standard demand systems cannot be estimated when aggregate zeroes are present in the data.
We explain how the problem arises in the workhorse discrete choice model introduced by McFadden
1973 and later extended by Berry et al. 1995 (which we refer to as BLP) for demand estimation.
When each product in the market has a separate fixed effect representing its mean utility in the
population of consumers, these fixed effects are estimated by equating the model’s prediction of
aggregate choice probability of each product with the empirical market share. Because the discrete
choice model never predicts aggregate zeroes, this moment condition cannot be solved when zero
demand is observed and estimation breaks down. However these mean utilities are the essential
ingredient used in BLP to estimate price elasticities.

A common approach to dealing with aggregate zeroes is to focus on the “top selling” products
in an industry which typically always have positive demand. However we show that this will
induce a selection problem that will bias demand elasticities towards zero. We show in Monte
Carlo simulations that this bias can be severe even with a relatively small fraction of zeroes in the
data (10-15 percent). The magnitude of the selection bias will be driven by the variance of the
unobservable in demand, which is exactly the role that retailer promotional effort plays in our model
as discussed above. This is critical for our question - if price elasticities are biased towards zero
standard techniques will overstate the importance of promotions in explaining demand fluctuations.
More generally demand studies that abstract away aggregate zeroes by selecting them away can
lead to severely downward biased demand estimates, and by implication upwardly biased estimates
markups, which will distort many policy questions.

Our main methodological contribution is to show how the standard discrete choice model for
demand can be estimated when there are aggregate zeroes in the data. As we show, zeroes are
direct indication that the number of consumers in each market is not sufficiently large relative to
the number of products/markets for consistency of the standard BLP estimator. However relaxing
their assumption introduces asymptotic bias in the estimates of demand. We solve the problem

2 We refer to these as market zeroes to distinguish them from corner solutions that appear in individual level
demand which was the main motivation for the development of discrete choice models (Domencich and McFadden
1975). In terms of an underlying discrete choice model an aggregate zero refers to a corner solution in demand that
persists at the market level after adding up over the individual discrete choices in the market.
by showing there is an optimal market share that minimizes this asymptotic bias. This optimal market share takes the empirical market shares off the boundary and thus handles zeroes in the data.

Our second key methodological result is that we show these optimal market shares can be easily computed under assumptions that are natural to the economics of the product differentiation. Demand in differentiated product markets often satisfy a form of Zipf’s law - there is a small number of products that receive the bulk of demand and a “long tail” of products with relatively thinner demand. This is clearly evident in our scanner data. We exploit a result from the statistics literature introduced by Chen [1980] to provide a foundation for Zipf’s law in terms of our demand model: when choice probabilities in a market are drawn from a Dirichlet distribution then quantities demanded in the market be distributed according to a long tail. We thus adopt the Dirichlet as an empirically useful approximation to the reduced form for the data generating process on choice probabilities in a market. The Dirichlet on the DGP for choice probabilities allows us to derive a closed form to an approximation of the optimal market share We show that this approximation works very well in Monte Carlos, which are the same ones used above to demonstrate the severity of the selection bias above.

We use our empirical strategy to revisit the test of the loss leader theory proposed by Chevalier et al. [2003] (hereafter CKR). CKR showed that the loss leader model makes a novel empirical prediction: the price of a product can actually fall if it becomes more popular because the incentive to promote it also become stronger. They used scanner data to show that the prices of products in categories with a seasonal demand shock fall, which is consistent with their theory. An especially stark effect they document is that the price index of tuna falls by 20 percent during Lent (which is a seasonal demand increase for the category). Nevo and Hatzitaskos [2006] (hereafter NH) casts doubt on loss leader being the explanation for this effect. They note that the average tuna price itself does not fall during Lent, and thus the fall in the price index can be explained by a change in the composition of purchases of tuna. They suggest the substitution towards cheaper varieties of tuna in the high demand period (Lent) can more likely be explained by demand becoming more elastic, which is consistent with alternative models of countercyclical markups (see e.g., Warner and Barsky [1995]).

Neither paper however studies scanner data at the disaggregate level and thus neither considers the role that sales prices are playing. We first show using their same data that the fall in the price of tuna during Lent is entirely due to the sales price index falling. This fall in the sales price index is due to a change in the composition of products that are purchased on sale, but there is no composition effect for products that are regularly priced. This is potentially consistent once more with the loss leader theory, but these same changes could also be explained by an increase in price sensitivity.

We first directly test the hypothesis that price elasticity is falling in the high demand period by applying our approach to demand estimation on the tuna data from DFF. We estimate demand using our optimal shares approach at the most disaggregate level in the data for both the high and
low demand periods. We find that controlling for products with zero demand using the optimal
shares (whose robustness we check against our moment inequality alternative) gives price elasticities
that are considerably more than twice as elastic than standard estimates of demand that exclude
the zeroes. Moreover we show that there is no evidence that demand is more price elastic in the
high demand period. Instead we show that the fall in the price index can be attributed to a change
in promotional strategies by the retailer - in the high demand period there is a tighter coordination
of discounting and promotion (both observed and more importantly for our purposes unobserved).
We view this as positive evidence in support of the loss leader model.

2 The Dominicks Scanner Data and Tuna Prices

We obtain data from Dominick’s Database through the Kilts center at the University of Chicago,
which covers weekly store-level scanner data at Dominick’s Finer Foods (DFF) and has been used
by many researchers The data comprises all Dominick’s Finer Foods chain stores in the Chicago
metropolitan area over the years from 1989 to 1997. Like other scanner data sets, this data set
provides information on demand at store/week/UPC level, where a UPC is a unique bar code that
identifies a product.

Chevalier et al. 2003 (which we refer to hereafter as CKR) put forward evidence in support
of the loss leader theory using the DFF data. We first replicate this evidence here, which in turn
motivates our empirical hypothesis. They showed that under loss leader pricing a product’s price
can actually fall after becoming more popular (a positive demand shock) because the incentive to
promote the product is also larger. Tuna during Lent is an especially telling comparison because this
seasonal demand shock is category specific and not associated with an overall increase in demand
for retail products.

CKR treat tuna as a product aggregate and measures its price using a price index. Let us index
weeks in the data by \( t \), stores by \( s \), and the UPC’s by \( j \). Let \( s_{jst} \) be the share of the total tuna sold
(in ounces) in week \( t \) of tuna \( j \) in store \( s \). Let \( p_{jst} \) be the corresponding price per ounce of tuna for

tuna \( j \) in store \( s \) at time \( t \). Then the price index for tuna in week \( t \) is

\[
P_t = \sum_{j,s} s_{jst} \log p_{jst}.
\]

The main result for tuna is shown in the first column in Table 1. This shows the coefficient on
the Lent indicator variable from a regression of \( P_t \) on \( Lent_t \) where the latter is an indicator that
takes value 1 if \( t \) is a Lent week. As can be seen the average price of tuna during lent appears to
be approximately 16 percent less expensive than other weeks. Thus the price of tuna significantly
falls during the category specific high demand period.

\footnote{For a complete list of papers using this data set, see the website of Dominick’s Database:
http://research.chicagobooth.edu/marketing/databases/dominicks/index.aspx}
Nevo and Hatzitaskos [2006] (which we refer to as NH hereafter) casts doubt on the loss leader theory being an explanation for this drop in price of tuna. They consider instead the average price of tuna, i.e.,

$$\bar{P}_t = \frac{1}{N_t} \sum_{j,s} \log p_{jst}.$$ 

where $N_t$ is the number of product/store pairs $(j, s)$ in the tuna data for week $t$. The second column of Table 1 shows the regression of $\bar{P}_t$ on $Lent_t$, which reveals that the average price of tuna is virtually unchanged between the Lent versus non-Lent period. NH rightly point out that the 20 percent reduction in the price of tuna is due to a change in the weights $s_{jst}$ between the Lent and Non-Lent periods, i.e., a composition effect. To understand the source of the change in composition we must consider the change in shares $s_{jst}$, which requires us to recognize that tuna is a differentiated product. If cheaper brands of tuna acquire higher market share in the high demand period because of changing tastes, i.e., more price sensitive demand, then there will be a substitution towards cheaper brands even without prices changing, which could explain the composition effect.

We will make NH’s point in a slightly different way that will prove instructive. We exploit the decomposition of a variable index into a “mean term” and a “reallocation” term that was introduced in the productivity literature by Olley and Pakes [1996]. Using this decomposition we have the simple accounting identity that

$$P_t = \bar{P}_t + \sum_{j,s} \Delta s_{jst} \Delta \log p_{jst}$$

which can be rewritten as

$$= \bar{P}_t + Cov_t$$

(2.1)

The “reallocation” or “covariance” term $Cov_t$ changes over time $t$ as the covariance between demand and price change. Thus all the change in the price index of tuna during Lent must be explained by the covariance changing during Lent, which we illustrate in the the third column in Table 1.

Explaining why tuna prices fall during the high demand period is now equivalent to explaining why the covariance term $Cov_t$ is dropping. NH suggest that a more plausible explanation than loss leadership is that demand is more elastic in the high demand (Lent) period. This increase in the elasticity will lead to a substitution towards cheaper products, which then will induce the covariance term to become more negative. Such an increase in elasticity could be due to search being intensive in the high demand period as modeled by Warner and Barsky [1995]. However if demand is becoming more elastic it would remain somewhat of a puzzle why the retailer is not
reducing the actual prices, i.e, the average price $\bar{P}_t$ should be falling.

NH make an important contribution in viewing tuna as a differentiated product and thus recognizing there can be substitution effects that drive changes in the price index of the product category. However the extent of product differentiation goes beyond there being different varieties of tuna. As discussed above, each UPC itself has two prices - a regular price and a sales price. The average price $P_t$ is thus a mixture of these two different prices. Thus it is entirely possible that the sales prices of products is falling while the regular prices are increasing and yet the average $\bar{P}_t$ stays the same.

We disaggregate the data a layer further than NH to explore this possibility. We consider the separate behavior of sales prices versus regular prices between the high demand and low demand period. We define a UPC $j$ at store $s$ on week $t$ as being on sale if it is either 5 percent below its highest price in the previous 3 weeks at the same store $s$. Because sales typically last for only one week this definition detects temporary sales and allows for both the regular price of UPC $j$ at store $s$ to change over time. In the first two columns of Table 2 below, we show the the regression of the price index for tuna on the Lent period separately for products on sale versus products marketed at their regular price. We see here a very distinct pattern - the price index of tuna falling during the lent period observed by CKR is entirely explained by the sales price index $P_{sale}^t$ falling, whereas regular prices have a very slight increase during Lent. We can decompose this change further by separately considering the average sales price $\bar{P}_{sale}^t$ and covariance term $Cov_{sale}^t$ from the decomposition of $P_{sale}^t$ according to (2.1). The results of these regressions are shown in remaining columns of Table 2. As can be seen the fall in the price of tuna is being driven completely by the fall in the covariance term $Cov_{sale}^t$ for products on sale. Thus the reason Tuna prices fall during lent is because there is a reallocation towards cheaper products that are on sale during lent. We observe no economically significant changes in either component for regular prices.

<p>| Table 2: Regression of Sales Price Index on Lent |
|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>$P_{sale}^t$</th>
<th>$P_{sale}^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{sale}^t$</td>
<td>-.20 (.03)</td>
<td>.01 (-.001)</td>
</tr>
<tr>
<td>$P_{regular}^t$</td>
<td>-.20 (.03)</td>
<td>.01 (-.001)</td>
</tr>
<tr>
<td>$Cov_{sale}^t$</td>
<td>-.001 (.0004)</td>
<td>(.0001)</td>
</tr>
<tr>
<td>$Cov_{regular}^t$</td>
<td>(.0001)</td>
<td>(.0001)</td>
</tr>
</tbody>
</table>

The change in $Cov_{sale}^t$ during the high demand period could potentially still be due to changing preferences. However it is difficult to reconcile why changing preferences would not also change covariance term of the regular price index. On the other hand the loss leader theory can explain this change in a consistent way. Suppose retailers more aggressively coordinate promotion and discounting in the high demand period, i.e., the correlation between the size of the promotional effort and the size of the discount is increasing in the high demand period. Then even without changing the sales prices themselves this change in promotion can induce a demand substitution

---

4 This is similar to the approach Hendel and Nevo 2006 take to defining a sale.
towards more steeply discounted products on sale. This would explain why the sales price index changes but the regular price index does not - promotional effort does not play a major role for products that are not discounted in the loss leader theory.

Is there any evidence directly in the data to suggest that the retailer might more tightly co-ordinates promotional effort and discounting in the high demand period. As already discussed in the introduction the full extent of promotional effort is likely to be unobserved. Nevertheless the data does provide some rough indication of promotion in the form of a deal variable - deal is an indicator that equals one if there is a flag for the product having a coupon or a feature display that week. It is known that for the DFF data there is a great deal of measurement error in this variable (see Glandon and Jaremski [2012] for a recent discussion), and even absent measurement error this is insufficient to capture the full extent of promotional effort. Nevertheless this variable might provide some suggestive evidence that a change in promotional behavior is indeed taking place between the two periods that is also consistent with the change in the $Cov_{t}^{sale}$ term.

In Table 3 we show the fraction of observations that are promoted according to the deal variable within the the different groups of sales price (a 5, 10, or 25 percent sale price) both in the high demand and low demand period. The results are consistent with the hypothesis that in the high demand period the correlation of promotional effort and discounting is increased - whereas the likelihood of having a deal does not change much between Non-Lent and Lent period for 5 percent or 10 percent discounts, it is much larger for 25 percent discounts and almost doubles for 50 percent discounts.

<table>
<thead>
<tr>
<th>Table 3: Observed Promotions and Discounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>5% Sale</td>
</tr>
<tr>
<td>Non-Lent</td>
</tr>
<tr>
<td>Lent</td>
</tr>
</tbody>
</table>

We now aim to test these differing explanations for why the price of tuna falls during lent using a structural model of demand facing the retailer. We seek to estimate the demand curve facing a retail store and test whether this demand is becoming more elastic in the high demand period and/or whether promotional effort is driving the reallocation effect in $Cov_{t}^{sale}$.

3 An Empirical Model of Retail Demand

3.1 Basic Setup

We now setup a model of demand for scanner data that provides the basis for addressing our empirical question. Each consumer entering the store makes shopping decisions as a function of the menu of price and promotion variables set by the store (which we shall refer to as the marketing variables). We will assume the retailer sets marketing variables on a weekly basis (this is the finest level of time aggregation we have available in the data). We focus on demand in a single product category that consists of different varieties of a differentiated product (such as tuna). Thus each marketing period at each store constitutes a separate market for the differentiated product.
will index markets by $t = 1, \ldots, T$ and let $J_t$ be the number of differentiated products in market $t$.

The shopping decisions among consumers at the store generates demand in the category which we measure as the market shares $s_t = (s_{1t}, \ldots, s_{J_t})$ of the $J_t$ products in market. We conceptualize this demand as arising from a demand function $s_t = D(p_t, x_t, \epsilon_t)$ where $p_t$ is a vector of the $J_t$ prices and $x_t$ is a vector of market level demand shifters (such as time trend, store effects, and demographic variables of the population of consumers at each store, measured characteristics of the products, observed measures of advertising, etc). The vector $\epsilon_t \in \mathbb{R}^{J_t}$ consists of unobservable demand shifters for each product. An important determinant of $\epsilon_t$ over the short time horizon will be unobserved promotional effort by the retailer, i.e., on a week to week basis at a given store promotions are varying simultaneously with price. Over a longer time horizon $\epsilon_t$ could capture changes in preferences for products.

One of the great strengths of scanner data is the sheer of markets that are available for empirical work - there are in principle close to 400,000 markets in our data that can be used to estimate $D$. Yet this wealth in the number of markets is balanced against the first key challenge in the data, which is that there is also a massive number of product varieties in each differentiated product category. Table 4 presents summaries of the extent of product differentiation for the categories in data. The first column shows the extent of products variety that is present in an average store/week - the number of UPC’s can be seen varying from roughly 50 (e.g., oatmeal and bath tissue) to over four hundred (e.g., cookies) within even these fairly narrowly defined categories. The second column shows that, within each category, there are relatively few “hit” products and a “long tail” of slow moving products (see Anderson [2007] for a general discussion of the long tail in markets). In particular, the data appears consistent with the well known “80/20” rule, which is a classic empirical regularity that finds roughly 80 percent of sales in a market being driven by the top 20 percent of products. The third column shows an important feature of the long tail - many of the slow moving products are often not sold at all in a given store/week market. In particular, we see that the fraction of observations with zero sales can even exceed 60% for some categories.
<table>
<thead>
<tr>
<th>Category</th>
<th>Average Number of UPC's in a Store/Week Pair</th>
<th>Percent of Total Sale of the Top 20% UPC's</th>
<th>Percent of Zero Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analgesics</td>
<td>224</td>
<td>80.12%</td>
<td>58.02%</td>
</tr>
<tr>
<td>Bath Soap</td>
<td>72</td>
<td>87.80%</td>
<td>74.51%</td>
</tr>
<tr>
<td>Beer</td>
<td>179</td>
<td>87.18%</td>
<td>50.45%</td>
</tr>
<tr>
<td>Bottled Juices</td>
<td>187</td>
<td>74.40%</td>
<td>29.87%</td>
</tr>
<tr>
<td>Cereals</td>
<td>212</td>
<td>72.08%</td>
<td>27.14%</td>
</tr>
<tr>
<td>Cheeses</td>
<td>283</td>
<td>80.41%</td>
<td>27.01%</td>
</tr>
<tr>
<td>Cigarettes</td>
<td>161</td>
<td>96.43%</td>
<td>66.21%</td>
</tr>
<tr>
<td>Cookies</td>
<td>407</td>
<td>81.40%</td>
<td>42.57%</td>
</tr>
<tr>
<td>Crackers</td>
<td>112</td>
<td>81.63%</td>
<td>37.33%</td>
</tr>
<tr>
<td>Canned Soup</td>
<td>218</td>
<td>76.25%</td>
<td>19.80%</td>
</tr>
<tr>
<td>Dish Detergent</td>
<td>115</td>
<td>69.04%</td>
<td>42.39%</td>
</tr>
<tr>
<td>Front-end-candies</td>
<td>199</td>
<td>76.24%</td>
<td>32.37%</td>
</tr>
<tr>
<td>Frozen Dinners</td>
<td>123</td>
<td>66.53%</td>
<td>38.32%</td>
</tr>
<tr>
<td>Frozen Entrees</td>
<td>346</td>
<td>74.65%</td>
<td>37.30%</td>
</tr>
<tr>
<td>Frozen Juices</td>
<td>94</td>
<td>75.16%</td>
<td>23.54%</td>
</tr>
<tr>
<td>Fabric Softeners</td>
<td>123</td>
<td>65.74%</td>
<td>43.74%</td>
</tr>
<tr>
<td>Grooming Products</td>
<td>461</td>
<td>80.04%</td>
<td>62.11%</td>
</tr>
<tr>
<td>Laundry Detergents</td>
<td>200</td>
<td>65.52%</td>
<td>50.46%</td>
</tr>
<tr>
<td>Oatmeal</td>
<td>51</td>
<td>71.37%</td>
<td>26.15%</td>
</tr>
<tr>
<td>Paper Towels</td>
<td>56</td>
<td>83.56%</td>
<td>48.27%</td>
</tr>
<tr>
<td>Refrigerated Juices</td>
<td>91</td>
<td>83.18%</td>
<td>27.83%</td>
</tr>
<tr>
<td>Soft Drinks</td>
<td>537</td>
<td>91.21%</td>
<td>38.54%</td>
</tr>
<tr>
<td>Shampoos</td>
<td>820</td>
<td>83.69%</td>
<td>69.23%</td>
</tr>
<tr>
<td>Snack Crackers</td>
<td>166</td>
<td>76.39%</td>
<td>34.53%</td>
</tr>
<tr>
<td>Soaps</td>
<td>140</td>
<td>77.26%</td>
<td>44.39%</td>
</tr>
<tr>
<td>Toothbrushes</td>
<td>137</td>
<td>73.69%</td>
<td>58.63%</td>
</tr>
<tr>
<td>Canned Tuna</td>
<td>118</td>
<td>82.74%</td>
<td>35.34%</td>
</tr>
<tr>
<td>Toothpastes</td>
<td>187</td>
<td>74.19%</td>
<td>51.93%</td>
</tr>
<tr>
<td>Bathroom Tissues</td>
<td>50</td>
<td>84.06%</td>
<td>28.14%</td>
</tr>
</tbody>
</table>

We can illustrate this long tail pattern in a different way by showing how the demand for a product decays by its rank in the market. Below we see the quantity sold by rank of product for bathroom tissues in a single store in a single week of our data, which exhibits a fast drop in demand as we move outside of the first few “hits” in the market, and a long tail of products with small relative demand. This exponential decay in demand by rank is often called Zipf’s law [Anderson, 2007]. The fact that the massive product variety in scanner data gives rise to the empirical regularity of
Zipf’s law within markets is an important point we will recall later.

Even the simplest linear demand system

\[ s_{jt} = \alpha_j p_{jt} + \beta_j p_{-jt} + \epsilon_{jt} \quad j = 1, \ldots, J \quad t = 1, \ldots, T \]

runs into a curse of dimensionality in this environment - with 200 products there will be 40000 parameters that are needed to identify the full pattern of own and cross elasticities. Once we allow demand parameters to vary by store and potentially evolve over time, even scanner data is too sparse to estimate a flexible demand system in the product space of SKU’s.

3.2 A Discrete Choice Approach

We appeal to the use of discrete choice models to deal with the curse of dimensionality in differentiated product demand. We illustrate the main ideas of our approach using the classic workhorse linear in parameters multinomial logit made famous by McFadden [1973]. We adopt this model for our presentation because it is both highly familiar in the literature and has just enough richness to speak to our problem. Yet even this seemingly textbook model encounters fundamental problems in our environment. All of the results we present below naturally generalize to other random utility models.

In discrete choice models each consumer \( i = 1, \ldots, N_t \) in market \( t \) makes a discrete choice
over the product varieties \( j = 0, \ldots, J_t \) where \( j = 0 \) denotes the outside option of not purchasing. Each product \( j \) is characterized by a price \( p_{jt} \) and bundle of hedonic attributes \( x_{jt} \in \mathbb{R}^{d_x} \) and an unobserved product attribute \( \xi_{jt} \). The observed product attributes will consist of SKU indicator variables, prices, and observed advertising. The unobserved product attribute \( \xi_{jt} \) will capture unobserved promotional effort and other demand changes. As mentioned above, if demand is relatively stable over short term horizon (such as the change from one week to the next) then \( \xi_{jt} \) is likely to mostly consist of unobserved promotional effort. Consumer \( i \) in market \( t \) is characterized by a vector of demographic variables \( z_i \in \mathbb{R}^{d_z} \).

In this setup the conditional indirect utility from consumer \( i \) purchasing product \( j \) in market \( t \) is

\[
 u_{ijt} = \delta_{jt} + z_{it} \Gamma x_{jt} + \epsilon_{ijt} \tag{3.1}
\]

where \( \Gamma \) is a \( d_z \times d_x \) matrix of parameters, \( \epsilon_{ijt} \overset{iid}{\sim} \text{Standard Type 1 Extreme Value} \), and

\[
 \delta_{jt} = \alpha p_{jt} + \gamma x_{jt} + \xi_{jt}. \tag{3.2}
\]

Thus the term \( z_{it} \Gamma x_{jt} \) captures all interactions between observed hedonic characteristics and consumer demographics and allows the demographics \( z_{it} \) to shift the marginal valuation of the characteristics \( x_{jt} \). The mean utility term \( \delta_{jt} \) plays the role of the “alternative specific constant” whose importance was emphasized by McFadden [1981] to control for unobserved product attributes. We note the fact that consumer demographics \( z_{it} \) only interact with the hedonic attributes \( x_{jt} \) and not price \( p_{jt} \) follows from an underlying assumption of quasi-linearity of utility in income (and thus marginal utility of income is constant across people), which is a convenient representation but can easily be relaxed.

The economics of product differentiation in the model are intuitive and provides a flexible representation of market demand. Observe that for a given \( z_{it} \) we can integrate out the vector of idiosyncratic taste shocks \( \epsilon_{it} = (\epsilon_{i0t}, \epsilon_{i1t}, \ldots, \epsilon_{iJ_t}) \in \mathbb{R}^{J_t+1} \) according to the extreme value distribution to yield the individual level choice probability of choosing \( j \in \{0, \ldots, J_t\} \) in market \( t \). Letting \( \nu_{ijt} = \delta_{jt} + z_{it} \Gamma x_{jt} \) denote the “representative” part of the random utility model (3.1), these choice probabilities take the logit form

\[
 \pi_{jt}(z_i) = \pi_{ijt} = \frac{e^{\nu_{ijt}}}{\sum_{k=0}^{J_t} e^{\nu_{ikt}}}
\]

where these logit choice probabilities vary by individuals \( z_i \) in the market. Integrating the logit choice probabilities over the population distribution \( F_t(z) \) of demographics in the market gives the market demand (i.e., the aggregate choice probability) for product \( j \in \{0, \ldots, J_t\} \)

\[
 \pi_{jt} = \int \pi_{jt}(z) \, dF_t(z) \quad j = 0, \ldots, J_t. \tag{3.3}
\]

\(^5\)In principle with microdata (i.e., individual choices within each market) the interaction parameters \( \Gamma_t \) can vary by market \( t \). We restrict \( \Gamma_t = \Gamma \) for convenience.
The choice probability system provides a flexible representation of aggregation demand. The cross price elasticity of demand \( \pi_{jt} \) with respect to \( p_{kt} \) using the standard properties of logit is

\[
\frac{\partial \pi_{jt}}{\partial p_{kt}} = -\alpha p_{kt} \pi_{jt} \hat{\pi}_{jt}(z) \pi_{kt}(z) dF(z)
\]

Thus the substitution patterns in the model are determined by the covariance effect

\[
\int \pi_{jt}(z) \pi_{kt}(z) dF(z)
\]

which is intuitive - when products \( j \) and \( k \) tend to be both popular and less popular for the same segments \( z \) of the population, they are interpreted as being closer substitutes to each other by the model. The magnitude of elasticities however also depend on the price parameter \( \alpha \).

### 3.3 Identification

The location normalization on the model is that \( \delta_{0t} = 0 \) for each market \( t \) and the scale normalization is the standardized variance of the logit errors \( \epsilon_{ijt} \). Identification of the model proceeds under two different sources of variation in the data. The first source of variation is the variation in consumers attributes and their choices, which was the key emphasis of the discrete choice literature following McFadden [1973]. We focus here on identification of the mean utilities \( \delta_T = (\delta_{11t}, \ldots, \delta_{J1t}, \delta_{12t}, \ldots, \delta_{J2t}) \) using this variation because they are the central issue for our problem (we thus treat \( \Gamma \) as known). Given the population distribution \( F_t \) of \( z \) in market \( t \) the demand system (3.3) can be represented as a system of equations in the mean utilities \( (\delta_{1t}, \ldots, \delta_{Jt}) \)

\[
\pi_{jt} = \sigma_t (\delta_{1t}, \ldots, \delta_{Jt}) \quad j = 1, \ldots, J_t
\]

which under general conditions (see Berry et al. [2013]) can be inverted

\[
\delta_{jt} = \sigma_t^{-1} (\pi_{1t}, \ldots, \pi_{Jt}) \quad j = 1, \ldots, J_t.
\]

Thus the aggregate choice probabilities \( \pi_t = (\pi_{1t}, \ldots, \pi_{Jt}) \) identify the mean utilities \( \delta_t = (\delta_{1t}, \ldots, \delta_{Jt}) \) through the inverse demand \( \sigma^{-1} \) implied by the model.

The second source of variation is at the level of products/markets. Using this market level variation to recover the parameters \( \beta = (\alpha, \gamma) \) that determine the mean utilities was the key innovation of Berry et al. [1995] (BLP for short) Because \( \delta_{jt} \) is identified for every product and every market from the choice data, the variation in products/markets allows us to identify the joint distribution of \( (\delta_{jt}, p_{jt}, x_{jt}, \tilde{z}_{jt}) \) where \( \tilde{z}_{jt} \) is an instrumental variable. The need for an instrument as emphasized by BLP is price \( p_{jt} \) is determined in market equilibrium and thus dependent in part on the unobservable \( \xi_{jt} \). In our data, given the assumption that \( \xi_{jt} \) reflects promotional effort the wholesale cost paid by the retailer serves as a natural instrument because wholesale cost is likely determined by long term contracts whereas promotions and discounts are more flexible strategic variables for the retailer. Indeed it has been documented that the pattern of price reductions in
scanner data is not reflected in the behavior of wholesale costs. On the other hand it is traditional
to assume that $E [\xi_{jt} | x_t] = 0$, i.e., unobserved characteristics are orthogonal to observed product
characteristics.

Letting $x_{jt} = (p_{jt}, w_{jt})$ and $z_{jt} = (x_{jt}, \bar{z}_{jt})$, we have that the parameters $\beta$ are identified by
standard instrumental variables

$$\beta' = E [z'_{jt} \delta_{jt}] (E [z'_{jt} x_{jt}])^{-1} \tag{3.5}$$

### 3.4 Estimation

The seemingly most natural approach to estimation is to use standard estimators correspond
to both stages of the identification of the model. The first step is to use maximum likelihood to
estimate the mean utilities $\delta_T = (\delta_{11}, \ldots, \delta_{1T}, \delta_{12}, \ldots, \delta_{JT})$ from the choice data, and then the
second step is to project the estimated fixed effects $\delta_{jt}^{MLE}$ onto the instrumental variables $z_{jt}$ to
estimate $\beta$ by two-stage least squares. This indeed corresponds precisely to the estimation algorithm
of BLP. Let us refer to the resulting estimator as $\beta^{BLP}$.

It is straightforward to show that $\delta_{jt}^{MLE} = (\delta_{1t}^{MLE}, \ldots, \delta_{JT}^{MLE})$ is the solution to a sample
analogue of the moment condition (3.4) that identifies $\delta_t$, i.e., MLE searches for the value of $\delta_t$ that
solves the sample moment restriction

$$s_{jt} = \int \pi_{jt} (z) d\hat{F}_t (z) = \hat{\sigma}_t (\delta_t) \tag{3.6}$$

where $\hat{F}_t$ is the empirical distribution of $z_i$ in market $t$ and $s_{jt}$ is the market share of product $j$ in
market $t$ in the sample. We will abstract from sampling variability in $z_i$ and treat $\hat{F}_t = F_t$. Thus
the resulting maximum likelihood estimator of the mean utilities becomes

$$\delta_t^{MLE} = \sigma^{-1} (s_t) \tag{3.7}$$

where $s_t = (s_{1t}, \ldots, s_{Jt})$ is the vector of empirical market shares. This system of equations (3.6)
is often called the “share equation” in the demand estimation literature.

### 3.5 The Zeroses Problem

The moment condition (3.6) now reveals the key problem we face. As we already saw in Table
the long tail pattern in the data causes many aggregate zeroes to arise in the data. The basic
problem that zeroes cause for the estimation of the model is that the system of equations (3.6)
cannot be solved if $s_{jt} = 0$ for any $j = 0, \ldots, J_t$. Thus $\delta_t^{MLE}$ is not defined. For example, in the
case of the simple logit where $\Gamma = 0$ the estimator (3.7) takes the simple closed form

$$\delta_{jt}^{MLE} = \log (s_{jt}) - \log (s_{0t})$$
and thus $\delta^{MLE}_{jt}$ clearly does not exist if $s_{jt} = 0$ as the log of zero is not defined.

### 3.6 The Selection Problem

This is straightforward to see that excluding observations with zero demand will lead to a selection bias. Consider the case where $E[\xi_{jt} | x_{jt}] = 0$. Then the standard BLP point estimates for the parameters $(\alpha, \beta)$ are equivalent to the OLS estimates of the model

$$\log \left( \frac{s_{jt}}{s_{0t}} \right) = \alpha_0 + \beta_0 x_{jt} + \xi_{jt}. \quad (3.8)$$

Clearly, the dependent variable is not defined when $s_{jt} = 0$ and thus those observations cannot be included in the formula of the point estimate. Thus the only way for the BLP estimator to be even defined for our data is to drop the observations with zero sales from the analysis. This is equivalent to redefining the outside good so that it subsumes all products with zero sales in a market.\(^6\) However this induces a selection problem that will bias the demand estimates.

The theoretical basis for this direction of the bias for the selection effect is clear. When $s_{jt} = 0$ for some $(j,t)$ and one drops these observations from the analysis, one is using a selected sample where the selection criterion is $s_{jt} > 0$. In this selected sample, the conditional mean of $\xi_{jt}$ is no longer zero. In fact,

$$E[\xi_{jt} | x_{jt}, s_{jt} > 0] > 0. \quad (3.9)$$

This is because the criterion $s_{jt} > 0$ selects high values of $\xi_{jt}$ and leaves out low values of $\xi_{jt}$. More importantly, the conditional expectation $E[\xi_{jt} | x_{jt}, s_{jt} > 0]$ is decreasing in $\beta_0 x_{jt}$. This is because for higher $\beta_0 x_{jt}$, the selection effect is smaller, and thus $E[\xi_{jt} | x_{jt}, s_{jt} > 0]$ is closer to zero (hence lower). Such a decreasing relationship implies that there will be attenuation bias if one simply estimate equation $(3.8)$ using the selected sample. The attenuation bias generally leads to demand estimators that appear to be too inelastic.\(^7\)

### 4 Our Solution: The Optimal Market Share

#### 4.1 The Basic Idea

On the face of it there seems an obvious solution to the zeroes problem. Instead of using the raw empirical share in $(3.7)$ we can use a different sample estimate $\hat{\pi}_{jt}$ of the aggregate choice probability $\pi_{jt}$ which does not take values on the boundary of the simplex. One possibility suggested a few

\(^6\)This is in fact the implicit strategy that standard statistical software would produce if one were to directly apply the BLP logit regression on the data, because the LHS variable $\ln \left( \frac{s_{jt}}{s_{0t}} \right)$ in the regression is not defined when $s_{jt} = 0$ and the default behavior is typically to drop such missing values from the sample that is used for computations. This selection of observations is also very similar to the common empirical strategy in the scanner data literature of focusing on the “top sellers” in a given product category when there is a large number of products, which implicitly defines all “long tail” products to be part of the outside good and hence they are selected out of the BLP regression.

\(^7\)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 are weakly dependent, the particular $t$ part of the selection dominates.
hundred years ago by the mathematician Laplace was

\[ \pi^{LS}_{jt} = \frac{N_t s_t + 1}{N_t + J_t + 1} \]

which corresponds to adding 1 to the demand of each product in the market (including the outside good \( j = 0 \)) and then forming new market shares based on these demands. Laplace motivated this construction from the point of view that even if an empirical frequency is zero we should never estimate an underlying binomial probability to be zero as this is a physical impossibility. He then derived \( \pi^{LS}_{jt} \) as an optimal Bayes estimator under a uniform prior on \( \pi_t \).

Any estimator \( \hat{\pi}_t \) that replaces \( s_t \) in the share equation (3.6) will lead to a consistent estimator \( \hat{\delta}_t \rightarrow p \delta_t \) as the number of consumers \( N_t \rightarrow \infty \) so long as \( \hat{\pi}_t \rightarrow p \pi_t \). The Laplacian \( \pi^{LS}_{jt} \) would thus be consistent (as the Bayes estimator and MLE both have the same limit). The question then becomes what is the right estimator \( \hat{\pi}_t \) of aggregate choice probabilities to use, i.e., what is the optimal way to take zeroes off the boundary. We now formalize the issue.

### 4.2 Introduction

Consider the product/market level identification equation:

\[ \sigma^{-1}_j(\pi_t) = x'_{jt} \beta + \xi_{jt}, \]

with an instrument vector \( z_{jt} \). For simplicity, assume that the model is just-identified, that is, \( x_{jt} \) and \( z_{jt} \) have the same dimension. In the ideal situation that \( \pi_t \) is observed, the estimation of \( \beta \) is straightforward. One can use the IV estimator:

\[ \hat{\beta}^* = \hat{Q}^{-1}_{xz} \hat{Q}_{zy}, \]

where \( \hat{Q}_{xz} = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} x_{jt} z'_{jt} \) and \( \hat{Q}_{zy} = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \sigma^{-1}_j(\pi_t) \). To show the consistency of \( \hat{\beta}^* \), observe that

\[ \hat{\beta}^* = \beta + \hat{Q}^{-1}_{xz} \hat{Q}_{z\xi}, \]

where \( \hat{Q}_{z\xi} = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \xi_{jt} \). Under the commonly used moment assumption that \( E \left[ J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \xi_{jt} \right] = 0 \) and that \( \{J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \xi_{jt}\}_{t=1}^{\infty} \) is stationary and weakly dependent, we apply a law of large numbers and show that, as \( n \rightarrow \infty \),

\[ \hat{Q}_{z\xi} \rightarrow_p 0. \]

Similarly, one can show that \( Q_{xz} \rightarrow Q_{xz} := E[ x_{jt} z'_{jt} ] \). Assuming that \( Q_{xz} \) is invertible, this implies the consistency of the ideal estimator \( \hat{\beta}^* \). In practical situations, \( \pi_t \) is not observed. Rather, the empirical market share \( s_t \) is observed instead. The empirical share \( s_t \) is computed by dividing the sales quantity vector \( q_t \) by the number of potential consumers \( n_t \). We assume that the consumers
are i.i.d., which implies that \( q_t \) is a draw from the multinomial distribution with parameters \((\pi_t, n_t)\).
The common practice (e.g. BLP) uses the plug-in estimator for \( \beta \):

\[
\hat{\beta}_{\text{BLP}} = \hat{Q}^{-1} \hat{Q}_z \hat{y},
\]

(4.5)

where \( \hat{Q}_z = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \sigma_j^{-1}(s_t) \). Similarly to (4.3), we have

\[
\hat{\beta}_{\text{BLP}} = \beta + \hat{Q}^{-1} \hat{Q}_z \xi + \hat{Q}^{-1} r_T,
\]

(4.6)

where \( r_T = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} z_{jt} (\sigma_j^{-1}(s_t) - \sigma_j^{-1}(\pi_t)) \). For \( \hat{\beta}_{\text{BLP}} \) to be consistent, it is necessary and sufficient that, as \( T \to \infty \),

\[
r_T \to_p 0.
\]

(4.7)

One set of sufficient conditions for (4.7) is given is Freyberger (2012). A key component of the sufficient conditions is that \( n_t \) increases with \( T \) sufficiently fast for each \( t \). While the asymptotic framework with (4.7) shows that \( \hat{\beta}_{\text{BLP}} \) is consistent, it does not necessarily imply that \( \hat{\beta}_{\text{BLP}} \) is a good estimator for \( \beta \) in finite samples. To see the finite sample problem clearly, consider a data set where \( s_{jt} = 0 \) for some \( j \) and \( t \). Then, for typical \( \sigma^{-1}(\cdot) \), assuming that \( z_{jt} > 0 \) for all \( j, t \), we have that

\[
r_T = -\infty \cdot 1_{d_z}.
\]

(4.8)

The minus infinity is clearly far from 0, indicating the asymptotic theory based on (4.7) is not a good approximation for the finite sample with \( s_{jt} = 0 \) for some \((j,t)\). In scanner data sets where each product is narrowly defined, it is the norm rather than exception that \( s_{jt} = 0 \) for some \((j,t)\). For such data sets, it is clear that \( \hat{\beta}_{\text{BLP}} \) is not a good estimator for \( \beta \).

### 4.3 Improved estimator

We aim to find an improved estimator for \( \beta \). The natural approach is to replace \( \hat{\pi}_t \) with a different estimator of \( \pi_t \). Because \( s_{jt} = 0 \) causes the extreme discrepancy between \( r_T \) and 0, we would like to the new estimator of \( \pi_t \) to avoid the boundary values. A simple choice is the Laplace transformation of \( s_t \):

\[
s_t^{lp} = (n_t + J_t + 1)^{-1}(n_t s_t + 1, d_t).
\]

(4.9)

The Laplace transformation adds a consumer to each product including the outside product and thus never hits the boundary of the probability simplex. Let \( r_T^{lp} \) be defined as \( r_T \) but with \( s_t^{lp} \) in place of \( s_t \). Then under the sufficient conditions in Freyberger (2012), \( r_T^{lp} \to_p 0 \) is satisfied, and such an asymptotic framework is no longer clearly inapplicable because \( r_T^{lp} \) is no longer a vector of \(-\infty\)'s. However, the Laplace transformation is just one of numerous possible transformations that can avoid the simplex boundary. For example, one could also add two consumers instead one to

\(^8\)Another common practice is to drop the \((j,t)\)s with \( s_{jt} = 0 \). But this causes sample selection bias because \( s_t \) is an endogeneous outcome variable.
each product. One could also simply add a small positive number to each \( s_{jt} \) that equals zero\(^9\). These transformations seem arbitrary and different transformations often lead to radically different estimators of \( \beta \). Here we propose an optimal transformation that minimizes a tight upper bound of the asymptotic mean squared error of the resulting \( \beta \) estimator. Let \( \hat{\pi}_t \) be a generic estimator of \( \pi_t \), constructed using observable information in market \( t \); that is, \( \hat{\pi}_t = \hat{\pi}(s_t, n_t) \) for a measurable function \( \hat{\pi} \). Note that \( s_{t}^{lp} \) as well as the transformations mentioned in the previous paragraph are special cases of \( \hat{\pi}_t \). Let the \( \beta \) estimator using \( \hat{\pi}_t \) be

\[
\hat{\beta}(\hat{\pi}) = \hat{Q}_{xz}^{-1} \hat{Q}_{zy}(\hat{\pi}),
\]

(4.10)

where \( \hat{Q}_{zy}(\hat{\pi}) = T^{-1} \sum_{t=1}^{T} J_t^{-1} \sum_{j=1}^{J_t} z_{jt} \sigma_j^{-1}(\hat{\pi}_t) \). We would like to study the mean squared error (MSE) of \( \hat{\beta}(\hat{\pi}) \) as an estimator of \( \beta \). The exact MSE is intractable. Thus, we would like to use some asymptotic approximation. For the effect of the estimation error in \( \hat{\pi}_t \) to remain in the asymptotic limit, we use an asymptotic framework different from Freyberger [2012]. In particular, we drop the assumption that \( n_t \) grows with \( T \) at a sufficiently fast rate (though we still allow that to happen). Then, by laws of large numbers and the continuous mapping theorem, we can show that, as \( T \to \infty \),

\[
\hat{\beta}(\hat{\pi}) - Q_{xz}^{-1} Q_{zy}(\hat{\pi}) \to_p 0,
\]

(4.11)

where \( Q_{zy}(\hat{\pi}) = E[z_{jt} \sigma_j^{-1}(\hat{\pi}_t)] \). In addition, by an appropriate central limit theorem and the Delta method, we can show that, as \( T \to \infty \),

\[
\sqrt{T}(\hat{\beta}(\hat{\pi}) - Q_{xz}^{-1} Q_{zy}(\hat{\pi})) \to_d N(0, V),
\]

(4.12)

for a variance-covariance matrix \( V \). Therefore, the asymptotic mean-squared error of \( \hat{\beta}(\hat{\pi}) \) is\(^{10}\)

\[
\text{Asy.MSE}(\hat{\beta}(\hat{\pi})) = \| Q_{xz}^{-1} E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))z_{jt}] \|^2 + \text{trace}(V)/T.
\]

(4.13)

Due to the scaling by \( 1/T \), the second summand on the right-hand side of (4.13) is typically of smaller order than the first summand. Thus, it is sufficient to focus on the first summand, which

\(^9\)When doing this, one implicitly deduct the small positive number from the empirical share of the outside good because the sum of all shares in a market should be one. Thus, the small number should be kept small enough so that the outside share does not become zero or negative after the deduction.

\(^{10}\)Because \( \hat{\beta} \) is a vector, we consider the sum of the MSE of each coordinate of it.
is the asymptotic bias part. Consider the following derivation for the asymptotic bias part:

$$\| \text{Asy.Bias}(\tilde{\beta}(\hat{\pi}_t)) \|^2 = \| Q_{xx}^{-1} E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t)) z_{jt}] \|^2$$

$$\leq E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2] E[z_{jt}^2 Q_{xx}^{-1} Q_{xx}^{-1} z_{jt}]$$

$$= C_0 E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2], \quad (4.16)$$

where $C_0$ is a positive constant. The bound is achieved when the inequality in (4.16) holds as an equality, that is when $(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))$ is perfectly correlated with each coordinate of $Q_{xx}^{-1} z_{jt}$. We thus would like to find a $\hat{\pi}$ such that $E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2]$ is minimized. To do so, observe that

$$E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2] = E\{E[(\sigma_j^{-1}(\hat{\pi}(s_t, n_t, J_t)) - \sigma_j^{-1}(\pi_t))^2|s_t, n_t]\}. \quad (4.17)$$

Thus, the optimal $\hat{\pi}(s_t, n_t, J_t)$ should be:

$$\hat{\pi}^*(s_t, n_t) = \arg \min_{a \in \mathbb{R}} E[(\sigma_j^{-1}(a) - \sigma_j^{-1}(\pi_t))^2|s_t, n_t]. \quad (4.18)$$

Because $\sigma^{-1} : \Delta^J \rightarrow R^J$ is a one-to-one mapping, the minimization problem in (4.18) has a closed-form solution:

$$\hat{\pi}^*(s_t, n_t) = \sigma \left( E[\sigma^{-1}(\pi_t)|s_t, n_t] \right). \quad (4.19)$$

### 4.4 Find the conditional distribution of $\pi_t$ given $(s_t, n_t)$

In practice, the conditional distribution of $\pi_t$ given $s_t, n_t$ (denoted $F_{\pi_t|s_t, n_t}$) is unknown. But this conditional distribution can be obtained if we know or can estimate two other quantities:

1. the conditional distribution of $s_t$ given $\pi_t$ and $n_t$, denoted $F_{s_t|\pi_t, n_t}$, and
2. the conditional distribution of $\pi_t$ given $(n_t, J_t)$, denoted $F_{\pi_t|n_t, J_t}$.

With $F_{s_t|\pi_t, n_t}$ and $F_{\pi_t|n_t, J_t}$, $F_{\pi_t|s_t, n_t}$ can be obtained through the Bayes rule:

$$F_{\pi_t|s_t, n_t}(p|s, n) = \frac{\int_x f_{s_t|\pi_t, n_t}(s|x, n) dF_{\pi_t|n_t, J_t}(x|n, J)}{\int_x f_{s_t|\pi_t, n_t}(s|x, n) dF_{\pi_t|n_t, J_t}(x|n, J)}, \quad (4.20)$$

Even when the second summand is not of smaller order, it may still be sufficient to focus on the first summand. To see this, suppose for simplicity that the markets are i.i.d. and $J_t = 1$ for all $t$. Then

$$V = Q_{xx}^{-1} \text{Var}(z_{jt}\sigma_j^{-1}(\hat{\pi}_t) - Q_{xy} z_{jt}) Q_{xx}^{-1}$$

$$\leq 2Q_{xx}^{-1} \left[ \text{Var}(E(z_{jt}\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))) + E(\xi_j^2 z_{jt} z_{jt}^t) \right] Q_{xx}^{-1}$$

$$= 2Q_{xx}^{-1} \left[ E(\xi_j^2 (\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2 z_{jt} z_{jt}^t) + E(\xi_j^2 z_{jt} z_{jt}^t) \right] Q_{xx}^{-1}. \quad (4.14)$$

Further assume that $z_{jt}$ is a bounded random vector. Then

$$\text{trace}(V) \leq C \left[ E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2] + 2E(\xi_j^2 z_{jt} Q_{xx}^{-1} z_{jt}^t) \right]$$

$$\leq C \left[ E[(\sigma_j^{-1}(\hat{\pi}_t) - \sigma_j^{-1}(\pi_t))^2] + 2E(\xi_j^2 z_{jt} Q_{xx}^{-1} z_{jt}^t) \right]. \quad (4.15)$$

The component that is affected by $\hat{\pi}_t$ is the same as the bound we derive for the asymptotic bias term in (4.13). Thus, it is without loss to focus on the asymptotic bias term in (4.13).
where \( f_{s|\pi, n_t} \) is the conditional probability mass function implied by the conditional distribution function \( F_{s|\pi, n_t} \). And then \( \hat{\pi}^* \) can be obtained accordingly. The conditional distribution \( F_{s|\pi, n_t} \) is known from the previous discussion: given \( \pi_t, n_t \), \( s_t \) is a draw from the multinomial distribution with parameter \((\pi_t, n_t)\), divided by \( n_t \):

\[
 n_t s_t \sim MN(\pi_t, n_t). \tag{4.21}
\]

The conditional distribution \( F_{\pi|n_t, J_t} \) is in general not known. However, we can infer about it from various sources of information. First, because the outside share \( s_{0t} \equiv 1 - s_1 J_t \) is never near either 1 or 0, for the size of \( n_t \) commonly seen in scanner data sets, it is safe to say that \( s_{0t} \) is an accurate estimator of \( \pi_{0t} \). Thus, we assume

\[
 \pi_{0t} \equiv 1 - \pi_1 J_t = s_{0t}. \tag{4.22}
\]

Second, we have already seen (recall Figure 3.1) that the histogram of the sales quantities in scanner data often can be described by the Zipf’s law; that is, \( \Pr(q_{jt} = q) \propto q^{-(1+\alpha)} \) for \( \alpha > 0 \) for large \( q \). Chen (1980) shows that this kind of quantity distribution can be generated if \( \pi_t/(1 - \pi_{0t}) \) follows a Dirichlet distribution. To describe Chen’s (1980) result, consider the simplified case that \( n_t \) and the share of the outside good \( \pi_{0t} \) are fixed at \( n \) and \( \pi_0 \), respectively. Then, Chen’s (1980) result can be stated in the proposition below in our notation.

**Proposition 1.** Suppose that \( \frac{\pi_t}{1 - \pi_{0t}} | J_t \sim Dir(\beta_1 J_t) \), where \( Dir(\beta_1 J_t) \) stands for the symmetric dirichlet distribution with parameter \( \beta_1 J_t \). Also suppose that \( \Pr(J_t/(n(1 - \pi_0)) \leq x) \to_d F(x) \propto x^\alpha \) for an \( \alpha > 0 \) as \( n \to \infty \). Then

\[
 E \left[ J_t^{-1} \sum_{j=1}^{J_t} 1(q_{jt} = q) \right] \to \phi(q) \propto q^{-(1-\alpha)}, \tag{4.23}
\]

where \( \phi(q) \propto q^{-(1-\alpha)} \) for large \( q \).

Because, under the assumptions of the proposition, the distribution of \( q_t \) given \( J_t \) is exchangeable across coordinates, we have

\[
 E \left[ J_t^{-1} \sum_{j=1}^{J_t} 1(q_{jt} = q) \right] = E \{ E \left[ J_t^{-1} \sum_{j=1}^{J_t} 1(q_{jt} = q) | J_t \right] \} = E \{ E[1(q_{jt} = q) | J_t] \} = \Pr(q_{jt} = q). \tag{4.24}
\]

Therefore, the proposition above establishes the connection between the Dirichlet distribution of \( \pi_t \) and the Zipf’s law for the sales quantities. In the general case that \( s_{0t} \) and \( n_t \) are not fixed, the connection holds conditional on \( n_t \) and \( \pi_{0t} \). Based on Chen’s (1980) result and the Zipf’s law
pattern of the sales quantities, we assume that
\[
\frac{\pi_t}{1 - \pi_{0t}} | J_t, n_t, \pi_{0t} \sim Dir(\theta 1_{J_t}),
\]
(4.25)
for an unknown parameter \( \theta \). This assumption combined with \( \pi_{0t} = s_{0t} \) gives us \( F_{\pi_t|s_{0t}, J_t} \) up to the unknown parameter \( \theta \). Even though \( \theta \) is unknown, it can be estimated through maximum likelihood.

By (4.21), (4.25) and \( \pi_{0t} = s_{0t} \), we have that
\[
\frac{s_t}{1 - s_{0t}} | J_t, n_t, s_{0t} \sim DCM(\theta 1_{J_t}, n_t(1 - s_{0t})),
\]
(4.26)
where \( DCM(b, n) \) stands for the Dirichlet compound multinomial distribution with parameters \( b \) and \( n \). Because \( s_t, J_t, n_t \) are observables, the parameter \( \theta \) can be estimated through standard maximum likelihood.

4.5 Practical guidelines and empirical Bayesian interpretation

Now we look back at \( \hat{\pi}^* \) in (4.19) and the conditional distribution \( F_{\pi|s_{0t}, n_t} \) in (4.20). Coincidentally, the Dirichlet assumption made in (4.25) guarantees a closed form solution for \( F_{\pi|s_{0t}, n_t} \), making the computation of \( \hat{\pi}^* \) much easier. It is well known that the Dirichlet distribution is a conjugate prior for the multinomial distribution, and that \( F_{\pi|s_{0t}, n_t} \), which can be viewed as the posterior distribution of \( \pi_t - s_{0t} \), is also Dirichlet:
\[
F_{\frac{s_t}{1 - s_{0t}}|s_{0t}, n_t} \sim Dir(\theta + n_t s_t).
\]
(4.27)

In practice, one can simply estimate \( \sigma^{-1}(\hat{\pi}^*_t) \) for each \( t \) by integrating \( \sigma^{-1}\left(\frac{s_t}{1 - s_{0t}}(1 - s_{0t})\right) \) under the above Dirichlet distribution with an estimated \( \theta \). Then, one can apply the \( \sigma(\cdot) \) mapping on the estimated \( \sigma^{-1}(\hat{\pi}^*_t) \) to obtain an estimator of \( \hat{\pi}^*_t \). This estimator can be viewed as an empirical Bayesian (EB) estimator of the choice probability \( \pi_t \). The computation can be even more simplified if we have a logit or nested logit model. In these cases, \( \sigma^{-1}(\hat{\pi}^*_t) \) can be obtained analytically. This is because, for logit and nested logit models, \( \sigma^{-1}\left(\frac{s_t}{1 - s_{0t}}(1 - s_{0t})\right) \) is composed of \( \log\left(\frac{s_t}{1 - s_{0t}}\right) \) and \( \log\left(\sum_{j \in J_t} \pi_{jt}/(1 - s_{0t})\right) \) for some sets of indices \( J \), and because for any random vector \( X := (X_1, \ldots, X_J) \sim Dir(\theta) \),
\[
E[\log(x_j)] = \psi(\theta_j) - \psi(\theta'1_{d_\theta}),
\]
(4.28)
where \( \psi(\cdot) \) is the digamma function, and
\[
E \left[ \log\left(\sum_{j \in J} x_j\right) \right] = \psi\left(\sum_{j \in J} \theta_j\right) - \psi(\theta'1_{d_\theta}).
\]
(4.29)
Table 5: Point Estimates for a Single Simulation

<table>
<thead>
<tr>
<th>Fraction of Zeros</th>
<th>16.70%</th>
<th>30.30%</th>
<th>44.80%</th>
<th>64.20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using Empirical Share</td>
<td>-.6178</td>
<td>-.3759</td>
<td>-.2517</td>
<td>-.0411</td>
</tr>
<tr>
<td>Using Laplace Rule</td>
<td>-.7323</td>
<td>-.5319</td>
<td>-.3402</td>
<td>-.1319</td>
</tr>
<tr>
<td>Inverse Demand EB</td>
<td>-1.0233</td>
<td>-1.0342</td>
<td>-1.0227</td>
<td>-1.2033</td>
</tr>
</tbody>
</table>

Note: T = 500, n = 10,000, True Value = −1.

5 Monte Carlo

In this section we present a Monte Carlo experiment that illustrates the performance of our optimal market share approach to handling zeroes in demand data. The Monte Carlo also shows the severity of the selection bias that can arise. We provide another Monte Carlo in the appendix that more closely resembles the features of our data. We consider here a binary choice logit model for ease of presentation.

Assume that each market has only one inside good and an outside good. The utility of consumer i in market t from consuming the inside good is \( u_{it} = \delta_t + v_{1it} \) and that from consuming the outside good is \( v_{0it} \), where \( v_{0it} \) and \( v_{1it} \) are independent type-I extremum value random variables that are independent of \( \delta_t \). Let \( \delta_t \) depend only on one single observed product characteristic \( x_t \) in the form \( \delta_t = \alpha_0 + \beta_0 x_t + \xi_t \). Then

\[
\log \left( \frac{\pi_t}{1 - \pi_t} \right) = \alpha_0 + \beta_0 x_t + \xi_t. \tag{5.1}
\]

For simplicity we take \( x_t \) and \( \xi_t \) to be independent so that \( x_t \) is its own instrument. Let the true value of the parameters be \( \alpha_0 = -1, \beta_0 = -1 \). We let \( x_t \sim Uniform[0, 15] \). Furthermore we let \( \xi_t \) exhibit a simple form of heteroskedasticity: \( \xi_t \sim N(0, 0.5^2) \times x_t \). Finally we vary \( \alpha_0 \) to produce different fractions of zeros in the simulated data. The number of consumer draws in each market that determine market shares \( s_t \) are taken to be 10,000, which is a very close to the number of consumers per market in our application.

Below (Table 5) we show the results for a single simulated sample and the point estimates corresponding to three possible constructions of market shares. Empirical shares refer to the standard use of the raw fraction of people who buy the product as the market share. Laplace shares refer to the Laplace’s “rule of succession” of adding one to the demand of each product to take shares off the boundary that was discussed above. Finally the optimal share refers to our calculation of the optimal share using a Dirichlet prior on the DGP for choice probabilities. As can be seen empirical shares and Laplace shares both suffer from a bias that attenuates the estimated value of the coefficient of interest (\( \beta_0 \)) towards zero. The optimal share by contrast stays remarkably close to the true value even when selection in the data is severe (over 60 percent of observations being zero).

In the next Table (Table we show a repeated simulation study based on 10,000 replications of the above design. We present the average bias for each previously estimator considered. Once again
### Table 6: Average Bias for a Repeated Simulation

<table>
<thead>
<tr>
<th>Fraction of Zeros</th>
<th>Using Empirical Share</th>
<th>Using Laplace Rule</th>
<th>Inverse Demand EB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16.48%</td>
<td>36.90%</td>
<td>49.19%</td>
</tr>
<tr>
<td></td>
<td>.3833</td>
<td>.6589</td>
<td>.7965</td>
</tr>
<tr>
<td></td>
<td>.2546</td>
<td>.5394</td>
<td>.6978</td>
</tr>
<tr>
<td></td>
<td>-.0798</td>
<td>-.0924</td>
<td>-.0066</td>
</tr>
</tbody>
</table>

Note: \( T = 500, n = 10,000, \) Number of Repetitions= 1,000.

we see the strong performance of the optimal share in environments even with severe selection.

### 6 Empirical Results for Tuna Demand

We now apply our empirical strategy to the data DFF on Tuna. Our goal is to examine whether the change in the composition of products purchased on sale, which we showed in Section 2 is the source of the fall in tuna prices during Lent, is due to demand becoming more elastic in the high demand period.

We employ the full sample of DFF data from 1989-1996 and all stores and estimate a single demand system. We employ a logit and nested logit model of demand. Under the logit model the product level regression is

\[
\log (s_{jt}) - \log (s_{0t}) = \alpha p_{jt} + \beta x_{jt} + \xi_{jt}
\]

where \( s_{0t} \) is the outside market share. We construct the outside share using the number of people who visited the store that week (the customer count) as the relevant market size. The control variables \( x_{jt} \) consist of UPC fixed effects and time trends (\( t \) and \( t^2 \)) along with the observed advertising (the deal variable discussed earlier).

We must deal with endogeneity of prices in the logit regression. 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.\(^{12}\) It is defensible in the disaggregated context we consider here because it has been shown that sales in retail price primarily reflect a reduction in retailer margins rather than a reduction in marginal costs (see e.g., Chevalier et al. [2003] and Hosken and Reiffen [2004]). Thus sales (and hence promotions) are not being driven by the manufacturer through temporary reduction in marginal costs. Nevertheless there do exist trade deals that cause variation in the marginal cost that may be correlated with unobserved promotion. We offset this possibility by using lags of marginal costs as further instruments.

In the nested logit specification we use the package sizes of the different UPC’s of tuna as the nest. We estimate the regression

\[
\log (s_{jt}) - \log (s_{0t}) = \alpha p_{jt} + \beta x_{jt} + \gamma \log (s_{jt|g}) + \xi_{jt}
\]

\(^{12}\) The gross margin is defined as (retail price - wholesale cost)/retail price, so we get wholesale cost using retail price \( \times \) (1 - gross margin).
where \( s_{jt} \) is the “within-group” share of product \( j \) in the market (its market share relative to the total group market share). The nesting parameter \( \gamma \) indicates the strength of correlation of individual tastes within a nest. If \( \gamma = 1 \) the correlation is perfect. If \( \gamma = 0 \) the correlation is zero and the model reduces to the simple logit. Nested logit introduces the further complication of constructing instruments for the additional endogenous variable \( s_{jtg} \). We use the number of products within the nest which exploits the fact there is entry and exit of UPC’s over time as well as a different selection of products at different stores.

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. The store level regressions turn are very similar to the pooled store regression and the latter is a more concise summary of demand behavior that we present here. The results are presented below in Table 7.

### Table 7: Demand Estimation Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Price Coefficient</th>
<th>Nesting Parameter</th>
<th>Average Own Price Elasticity</th>
<th>Fraction of Inelastic Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit Emp. Shares</td>
<td>-.51 (.&lt;.01)</td>
<td>-</td>
<td>-.77</td>
<td>82.82 %</td>
</tr>
<tr>
<td>Opt. Shares</td>
<td>-2.01 (.01)</td>
<td>-</td>
<td>-3.01</td>
<td>.33 %</td>
</tr>
<tr>
<td>Nested Logit Emp. Shares</td>
<td>-.52 (.&lt;.01)</td>
<td>.51 (&lt;.01)</td>
<td>-1.50</td>
<td>29.26 %</td>
</tr>
<tr>
<td>Opt. Shares</td>
<td>-.98 (&lt;.01)</td>
<td>.82 (&lt;.01)</td>
<td>-7.56</td>
<td>&lt;.01 %</td>
</tr>
</tbody>
</table>

Note: The instrumental variables for price include wholesale price, its first and second lags (for the same product/store). IV for the within group (nest) share is the number of products in the group.

Our main results are relatively clear from the table. The first main result is that using the standard estimator in the literature that inverts empirical shares to recover mean utilities suffers from serious selection bias towards zero. The UPC level elasticities for the logit model are unrealistically small, with the average elasticity in the data being -.77. Furthermore, over 80 percent of products having inelastic demand. Using our optimal shares instead of empirical shares has a major effect. Demand for UPC’s starts to look realistically elastic - the average elasticity being -3 and less than 1 percent of observations having inelastic demand.

The severity of the inelastic demand is lessened with empirical shares when moving to nested logit - the average elasticity doubles to -1.50 and the number of observations with inelastic demand falls to less than 30 percent - reflecting the importance of the nesting parameter. However demand at the UPC level remains relatively rather inelastic. The demand estimates using the optimal shares shows a similar contrast between nested logit and logit. In particular, the nesting parameter becomes grows larger under the larger and demand elasticities on average double relative to logit. The number of products exhibiting inelastic demand falls to under .1 percent. However the key
remains that demand is considerably more elastic with the optimal market share. We believe these are the first results in the literature using fully disaggregated scanner data that generates realistically large estimates of own elasticities at the UPC level (where there are over 100 varieties of tuna that should be fairly substitutable).

Our second main results is that we do not find evidence to suggest demand is becoming more elastic in the high demand period. These results are shown in Table 8. If anything, demand appears more inelastic during Lent. This is amplified by the nested logit parameter primarily because the strength of package size preferences are not as strong during Lent. The actual price sensitivity appears very similar across the two periods. Using empirical shares rather than the optimal shares under logit one may falsely conclude that price sensitivity is increasing in the high demand period. This stresses again the applied importance of the optimal market shares.

<table>
<thead>
<tr>
<th>Table 8: Demand in Lent vs. Non-Lent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Coefficient Nesting Parameter Average Own Price Elasticity</td>
</tr>
<tr>
<td>Logit</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Opt. Share</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Nested Logit</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Opt. Share</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: The instrumental variables for price include wholesale price, its first and second lags (for the same product/store). IV for the within group (nest) share is the number of products in the group.

If demand is not becoming more elastic, what then explains the movement in the $Cov_{it}^{sale}$ term in the high demand period? Recall that our alternative explanation is that there is a tighter coordination of promotional effort and discounting in the high demand period. This was consistent with our observable measure of promotion - we saw in Table 3 that for steeper discounts there was a greater likelihood of being placed on deal in the high demand period. We see the same pattern arise with unobserved promotional effort. The correlation between $p_{jt}$ and $\xi_{jt}$ among products that are flagged as being on sale (having at least a 5 percent discount relative to the regular price) doubles from -.08 to -.16 between the Non-Lent and Lent periods. This is consistent with a stronger correlation of promotional effort (captured through $\xi_{jt}$) and discounts $p_{jt}$ in the high demand period.

To make this point a different way and give an order of magnitude appreciation for the size of the unobserved promotion effort $\xi_{jt}$ in the estimates, we consider the following counterfactual. We counterfactually change all sale prices in the data to the regular price based on the definition of sales and regular price discussed in Section 2. We let the residual $\xi_{jt}$ equal to the implied residual from
the demand model (using the nested logit specification over the whole sample) from the actual data (i.e., from the “factual”). We then predict new counterfactual demands $s_{jt}^{\text{predict}}$ using the model. We then reconstruct the price index of Tuna explained in Section 2 using these counterfactual prices and quantities. We then study the change in tuna prices during Lent for this counterfactual price index. The results are shown in Table 9. Remarkably the price of tuna falling during Lent (originally documented in Table 1) can be largely explained by the change in promotional effort in the high demand period alone.

<table>
<thead>
<tr>
<th></th>
<th>P (Price Index)</th>
<th>P (Average Price)</th>
<th>Cov (Composition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lent</td>
<td>-.14</td>
<td>-.02</td>
<td>-.12</td>
</tr>
<tr>
<td>s.e.</td>
<td>(.0005)</td>
<td>(.0003)</td>
<td>(.0005)</td>
</tr>
</tbody>
</table>

7 Conclusion

(TO BE COMPLETED)

8 Appendix: Monte Carlo with Many Products

We run a set of Monte Carlo experiments based on nested logit model. Let $J_t = 50$ and $n_t = 15,000$ for $t = 1, \ldots, 100$. Normalizing the mean utility of outside option to zero, we define the mean utility of inside goods as

$$\delta_{jt} = c_0 + \beta_0 x_{jt} + \xi_{jt}.$$ 

For simplicity, we consider a simple nesting structure that has three groups: the first and last 25 inside goods are in two nests respectively; the outside option itself is in one nest. If product $j$ is in group $g$, then the market share of $j$ is

$$s_{jt} = \frac{\exp \left[ \delta_{jt}/(1-\lambda) \right]}{(D_g)^\lambda \left[ \sum_{g'} (D_{g'})^{1-\lambda} \right]},$$

where $D_g \equiv \sum_{j \in g} \exp \left[ \delta_{jt}/(1-\lambda) \right]$ and $\lambda$ is the nesting parameter that is assumed to be the same across groups.

The remaining data generating process is as follows. For each market $t$, let $x_{jt} = \frac{j}{10} + \text{Uniform}[0, 1]$ and $\xi_{jt} = [3 \cdot \exp (10 - x_{jt})/R] \cdot N(0, 1)$, where $R$ is the range of $\exp (10 - x_{jt})$ in the simulated data. Note that $x_{jt}$ has some persistence across markets - products with larger index tend to have higher value of $x$. We vary $c_0$ to get different fractions of zero shares in the generated data.

We compare our optimal share (which we label below as EB for empirical Bayes) estimator with two alternatives. One is the BLP estimator that uses the empirical shares and drops the observations.
with zero share. The other one is the simple Laplace estimator that modifies the empirical shares by adding 1 purchase to all the products. Since the within group share is endogenous, we use the sum of \( x \) within the group, \( \sum_{j \in g} x_{jt} \), as the instrument for within group share. Table 10 reports the estimation results.

<table>
<thead>
<tr>
<th>( c_0 = -8 )</th>
<th>( \beta )</th>
<th>( \lambda )</th>
<th>Fraction of Zeros</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLP</td>
<td>Laplace</td>
<td>EB</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>1.27</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.023)</td>
<td>(.010)</td>
</tr>
<tr>
<td>( c_0 = -9 )</td>
<td>.80</td>
<td>1.33</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.026)</td>
<td>(.010)</td>
</tr>
<tr>
<td>( c_0 = -10 )</td>
<td>.72</td>
<td>1.36</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>(.007)</td>
<td>(.029)</td>
<td>(.011)</td>
</tr>
<tr>
<td>( c_0 = -11 )</td>
<td>.62</td>
<td>1.28</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.026)</td>
<td>(.011)</td>
</tr>
</tbody>
</table>

Note: True parameter values are \( \beta_0 = 1 \) and \( \lambda_0 = .5 \).

References


