PRODUCT VARIETY, ACROSS-MARKET DEMAND HETEROEGENEITY, AND THE VALUE OF ONLINE RETAIL

Thomas W. Quan
Department of Economics
University of Georgia

Kevin R. Williams*
School of Management
Yale University

September 2015

Abstract

This paper quantifies the effect of increased product variety in online markets on consumer welfare and firm profitability. While online retail gives consumers access to an astonishing variety of products, the value of this variety depends on the extent to which local demand can be captured by local retailers. We use a large, detailed data set from a popular online retailer. However, despite observing millions of transactions, many products have zero market share at the local level. We propose a modification to Berry (1994) and Berry, Levinsohn, Pakes (1995) that addresses this sampling problem. Our estimation technique is easy to implement and fits the data well. Our results indicate that products face substantial heterogeneity in demand across geographic markets which greatly mitigates the consumer gains of increasing variety when traditional brick-and-mortar retailers localize assortments. This suggests the long revenue tail found in online retail may be the consequence of demand aggregation, where aggregating over local markets with differing demand creates the appearance of large within market heterogeneity at the national level.

JEL: C13, L67, L81

*Authors: quan@uga.edu and kevin.williams@yale.edu. We are very grateful to the anonymous online retailer for allowing us to gather the data set used in this paper. We thank Amit Gandhi, Thomas Holmes, Kyoo il Kim, Amil Petrin, Catherine Tucker, Joel Waldfogel, and the many seminar and conference participants for useful comments. We also thank the Minnesota Supercomputing Institute (MSI) for providing computational resources and the online retailer for providing us with data.

1 Introduction

There is widespread recognition that as economies have advanced, consumers have benefited from an increasing access to variety. Several strands of the economics literature have examined the value of new products and increases in variety either theoretically or empirically, e.g. in trade (Krugman 1979), macroeconomics (Romer 1994), and industrial organization (Lancaster 1966, Dixit and Stiglitz 1977). The internet has given consumers access to an astonishing level of variety. Consider shoe retail. A large traditional brick-and-mortar shoe retailer offers at most a few thousand distinct varieties of shoes. However, as we will see, an online retailer may offer over 50,000 distinct varieties. How does such dramatic increases in variety contribute to welfare?

The central idea of this paper is that gains from online retail depend critically on the extent to which demand varies across geography and how traditional brick-and-mortar stores respond to local tastes (Waldfogel 2010).[1] For example, the addition of 5,000 different kinds of winter boots online will be of little value to consumers living in Florida just as the addition of 5,000 different kinds of sandals online will be of little consequence to consumers in Alaska. If Alaskan retailers offer a large selection of boots that capture the majority of local demand, only consumers with niche tastes – possibly those who want sandals – will benefit from the variety offered by online retail. Therefore, in order to quantify the gains from variety due to online retail, it is critical to estimate the extent to which demand varies across regions.

We have collected an extremely detailed data set consisting of point-of-sale, product review, and inventory data from a large online retailer. One of the product categories the retailer sells is footwear, and we observe over 13.5 million shoe sales across thousands

[1] Recent literature has highlighted heterogeneity in demand across markets. For example, in a series of papers, Waldfogel finds evidence of differences in demand across demographic groups in radio (Waldfogel 2003), television (Waldfogel 2004), and chain restaurants (Waldfogel 2008). Bronnenberg, Dhar, and Dube (2009) document a persistent early entry effect on a brand's market shares and perceived quality with stronger effects in markets geographically closer to a brand's city of origin. Finally, Bronnenberg, Dube, and Gentzkow (2012) find that brand preferences can explain 40% of geographic variation in market shares.

of products. In addition, we collected data on shoe assortments across stores for a few large retail chains. This data provides us with direct evidence that firms are responding to across-market heterogeneity, as product assortments vary significantly across stores. With our transactions data, we document large differences in demand for specific products across geographic markets. Since prices, product characteristics, and choice sets are the same for all geographic markets, these differences can only be rationalized by differences in local demand. To formally test for differences across markets, we use simple multinomial tests that overwhelmingly reject the null hypothesis that consumers across markets have the same demand over shoes.

After showing that the data is inconsistent with a model devoid of across-market demand heterogeneity, we turn to estimating the gains from online variety. Our modeling approach closely follows the discrete choice literature with an emphasis on allowing tastes to vary across locations. The importance of flexibly modeling heterogeneity in discrete choice setups has been well documented in the literature (see Berry, Levinsohn, and Pakes (1995), Petrin (2002), Song (2007)). The model allows for the fact that, for example, the removal of a popular sandal will be much more costly for markets in Florida than for markets in Alaska.

Employing our data at the level of narrowly defined products and at narrow geographic detail, however, also presents us with an empirical challenge. Despite the fact that we observe millions of transactions, the large number of products and locations inevitably leads to many products having local market shares equal to zero. For example, 59.8% of products have zero sales at the annual-state level.[2] The zeros are problematic for two reasons. First, using standard demand techniques creates a selection bias in the demand estimates (Berry, Linton, and Pakes 2004, Gandhi, Lu, and Shi 2013, Gandhi, Lu, and Shi 2014), leading to incorrect estimates of the gains of increased variety. Second, the zeros suggests a small samples problem. This is particularly problematic for us because if

[2]Note that aggregation over the time horizon is also problematic because of the high turn over in products. Conlon and Mortimer (2013) highlight that ignoring these changes to the choice set may bias demand estimates.

uncorrected, we would overstate the degree of heterogeneity across markets (Ellison and Glaeser (1997)) and understate the gains of increasing variety. For example, suppose on a particular day, a market sees a single shoe sale and that is the only sale of that product. If we fail to account for the small samples issue, we would come to the conclusion the rest of the country has absolutely no interest in the product, and that the market would not benefit from more variety because consumers only want that shoe.

A contribution of our paper is to propose solutions to these issues in the discrete choice demand framework. Our estimation strategy exploits the structure of the demand model to separate the problem into two parts. At the aggregate level, our approach effectively mimics the standard approach and we are able to pin down the price coefficient and other parameters common across markets. We bring in the local market share information to form a set of micro moments that augment the aggregated (national) sales data (Petrin 2002) while explicitly accounting for small samples. Our approach allows us to identify the variance of product-market level random effects, instead of recovering product-market level fixed effects.

Our results indicate that demand for specific products vary significantly across markets.[3] We show that accounting for this heterogeneity is necessary for rationalizing the distribution of local sales. When brick-and-mortar retailers cater their assortments to local demand, we find that the welfare gains from online variety are relatively small. About 6.8% of the total unconstrained consumer welfare is due to online variety. However, if we shut down the across-market demand heterogeneity, and hence the localization in brick-and-mortar retail, we would find 19.5% of the unconstrained consumer welfare is due to online variety, an overstatement of 187%. Put another way, if local stores cater to the local demand, then the value of online markets is relatively small because the average consumer already has access to the products they want to purchase.[4] Additionally, for

[3]We estimate the extent to which demand differs across locations, but not causes for differences. Some natural reasons would be weather and local fads. Tucker and Zhang (2011) look at how popularity affects mainstream vs. niche products.

[4]Our data does not allow us to model competition between retail and our online retailer. Brynjolfsson,

brick-and-mortar retailers, we find a large incentive for them to cater to their local demand. By doing so they can obtain 14.5% higher revenue than under a standardized assortment.

Finally, our results suggest a new interpretation of the long tail phenomenon of online retail (Anderson 2004). The term describes a shift in the distribution of revenue toward niche, or tail, products. The prevailing view is that the long tail pattern has emerged because niche products better satisfy consumer tastes, leading consumers to switch from purchasing hit products available at their local brick-and-mortar retailers to purchasing niche products only available online.[5][6] This suggests sizable welfare gains from increasing access to variety. However, we find relatively smaller gains from increasing access to variety for shoes. This is due to the fact that in the presence of across-market demand heterogeneity, aggregating demands across markets yields a long tail nationally, but at the local level, the tails remain relatively short, which traditional brick-and-mortar retailers can capture through localization of assortments.[7] This suggests that the long-tail may be driven by aggregating across markets with differing demand, instead of very large differences in demands within markets.

The rest of the paper is organized as follows. Section 2 discusses our data and presents preliminary evidence of across-market heterogeneity. In section 3, we present the model and estimation procedure. Results and counterfactuals are in Section 4 and 5, respectively.

Hu, and Rahman (2009) finds that competition is higher for mainstream products, and online retailers face less competition for niche products.

[5] A counterpoint can be found in Tan and Netessine (2009). They use individual level data on online movie rentals and find no evidence that niche titles satisfy consumer tastes better than hit titles. Instead niche consumption is driven by a small subset of heavy users. Additionally, they find a shortening effect on the tail with the addition of new products. They conclude that this is due to new titles appearing faster than consumers can discover them.

[6] It has been suggested that these gains may be increasing over time as papers using multiple years of data have found the long tail to be getting longer. (Chellappa, Konsynski, Sambamurthy, and Shivendu 2007, Brynjolfsson, Hu, and Smith 2010).

[7] To see this consider the following example: Suppose there are 100 equally sized markets, and each prefers a different good. In each market, the local brick-and-mortar retailer sells one good that makes up 100% sales (short tail). Now suppose an online retailer enters, which gives all 100 markets access to all 100 products. Assuming an equal number of consumers from each market purchase online, the online retailer will sell 100 goods that each make up 1% of sales (long tail). In this example the welfare gain from access to variety would be zero, since all consumers were already being served their preferred good by their local brick-and-mortar.

Section 6 discusses the robustness of our findings and Section 7 concludes the paper.

2 Data

We create several original data sets for this study. The main data set consists of detailed point-of-sale, product review, and inventory data that we collected from a large online retailer. With this data, we observe over $1 billion worth of online shoe transactions between 2012 and 2013. We augment this with a snapshot of shoe availability for a few large brick-and-mortar retailers.

We begin by summarizing our data sets (Section 2.1). We then provide evidence of across market heterogeneity using the brick-and-mortar assortment data (Section 2.2) and with the online retail data (Section 2.3). Finally, we document the small samples problem in the data – in particular, the zeros problem – and discuss aggregation as a means to address the issue (Section 2.4).

2.1 Data Summary

Online Retailer Data

The main data set for this study was collected and compiled with permission from a large online retailer. This online retailer sells a wide variety of product categories, including footwear, which will be the focus of our analysis. Each transaction in the point-of-sale (POS) data base contains the timestamp of the sale, the 5-digit shipping zip code, price paid, and a wealth of information about the shoe, including model and style information. The transaction identifier allows us to see if a customer purchased more than a single pair of shoes. Finally, we download a picture of each shoe, and image process them to create color covariates.

We merge in product review and inventory data. The review data contains the time series of reviews and ratings for each shoe. For the inventory data, we track daily inventory for every shoe.[8] Importantly, this data allows us to infer the complete set of shoes in the consumer's choice set, even when the sale of a particular shoe is not observed. While the inventory data is size specific, the sales data does not include size. We concede this in general will cause us to understate the gains of online variety because consumers with unusual foot sizes greatly benefit from online shopping since traditional retailers may not typically stock unusual sizes.[9]

We observe over 13.5 million shoe transactions during the collection period, with a majority of transactions being women's shoes. The price of shoes varies substantially across gender, but also within gender – for example, dress shoes tend to be more expensive than walking shoes. The distribution of transaction size per order is heavily skewed to the left. Only a small fraction of orders contain several pairs of shoes. Additionally, of the transactions containing multiple purchases, less than a quarter contain the same shoe, suggesting concern over resellers is negligible in our data set. This also implies there are few consumers buying multiple sizes of the same shoe in a single transaction. Overall, we believe this supports our decision to model consumers as solving a discrete choice problem.

We observe over 580,000 reviews of products and record the consumer response to a few questions regarding the fit and look of the product. The metrics we include in the demand system are ratings for comfort, look, and overall appeal, where 1 is the lowest rating, and 5 is the highest rating. The reviews are heavily skewed towards favorable ratings.

An important feature of the data is the number of products the online retailer offers. The average daily assortment size is over 50,000 products, and over the span of data collection, over 100,000 varieties of shoes were offered for sale. This constantly changing

[8]Initially this data was not collected daily, but for the last seven months of data collection, each shoe inventory was tracked daily.

[9]However, one store manager we spoke to indicated his retailer sets assortments based on not only styles, but also on sizes. With our brick-and-mortar data, we can test for this. With our Macy's data, we can reject the null hypothesis that the mean assortment size is constant across stores.

choice set provides us with additional variation that will help us identify the parameters of our model.

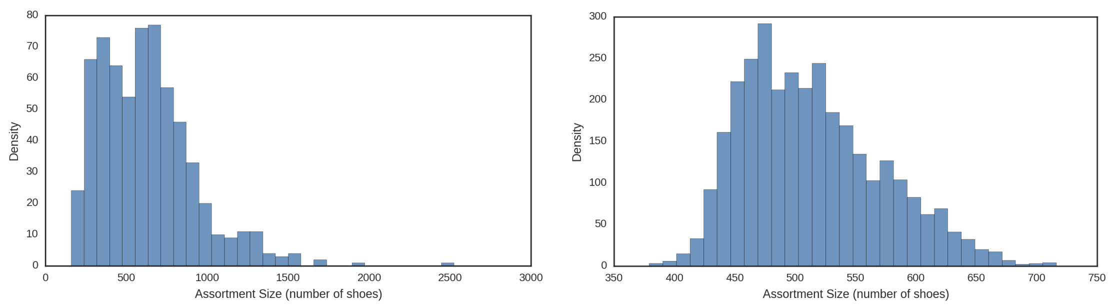
Brick-And-Mortar Data

In addition to the retail data, we collect a snapshot of shoe availability from Macy's and Payless ShoeSource during August and September of 2014. We first collected all the shoe SKUs each retailer sold, and then for each SKU, we used the firm's "check in store" web feature to see if the product was currently available. The firms' websites do not list how many shoes are in stock, just whether a shoe is available or not. Since each query was for a specific shoe size, we then aggregate across all sizes to have a measure of product availability. This allows us to compare assortments with our online data and corrects for the possibility that a particular size may be sold out at a given store. If across-market consumer demand heterogeneity is as important as we claim, we would expect to see brick-and-mortar retailing chains stocking different products at different locations. Assortment data from Macy's and Payless provide clear evidence of this.

Table 1: Summary of Brick-and-Mortar Data

	Macy's	Payless Shoes
Number of stores	649	3,141
Number of products	7,844	1,430
Percent online exclusive	34.8%	19.2%
Avg. assortment size	624.9	513.0
	(299.3)	(58.4)

Table 1 presents summary information on Macy's and Payless' assortments. In September 2014, we observe 7,844 different styles available at Macys.com. About 35% of shoes are online exclusives, making just over 5,000 shoes available at least one of 649 physical locations. At Payless.com, we observe 1,430 distinct styles, with about 19% being online exclusives. Average in-store assortment sizes are similar across retail chains - 624.9 and 513 for Macy's and Payless, respectively. However, there is much greater variance in Macy's store size. Figure 1 highlights these differences in the form of histograms of the assortment sizes at Macy's and Payless locations. Unsurprisingly, we find that the stores with larger assortments tend to be located around larger population centers.



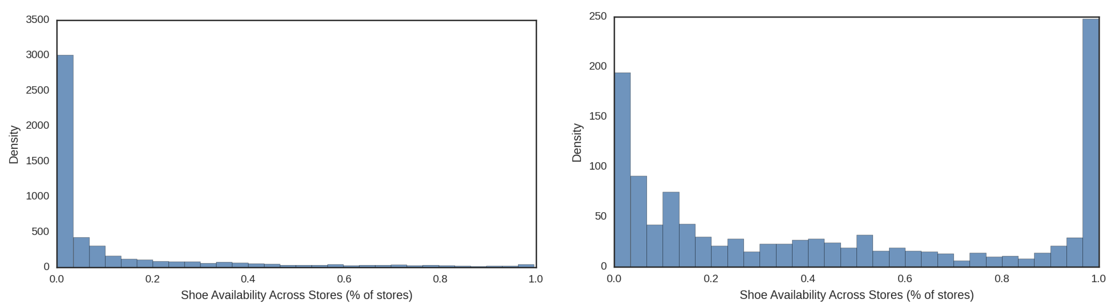
(a) Macy's (b) Payless

Figure 1: Histogram of shoe assortment size across stores

2.2 Localization of Brick-And-Mortar Retailers

The premise of this paper is that there may exist significant differences in consumer demand across geographic markets. This has significant implications for the value of increased variety if traditional brick-and-morter retailers localize assortments. For large national retailers, there are trade offs to localizing assortments. On the one hand, catering to local demand can greatly increase revenues, but on the other hand, there are advantages from economies of scale through standardization. In this section, we present evidence from the product assortment choices of Macy's and Payless. While there could be supply-side reasons for localized assortments, such as different transportation costs across brands, we find this to be an unlikely rationale, since the vast majority of products are produced overseas. We interpret the facts below as evidence of firms responding to demand.

We want to measure how assortments vary by store. Figure 2 graphs the percentage of locations carrying a shoe style for Macy's and Payless. If all shoes were available at all stores, the density would collapse at 1 (100%). The analysis excludes online only shoes. For Macy's we can see that the vast majority of products are sold at only a few stores; that is, the density is concentrated primarily to the left. The Payless Shoes distribution is more bimodal, at few stores and at almost all stores. In recent years, Macy's has made a concerted effort to better localize their product assortments through a program called "My Macy's."[10] The strikingly low prevalence of products across stores is likely reflective of this program. Payless, on the other hand, produces and partners with other brands to provide exclusive products for its retail chain. The bimodal distribution for Payless may be reflective of these partnerships.



(a) Macy's (b) Payless

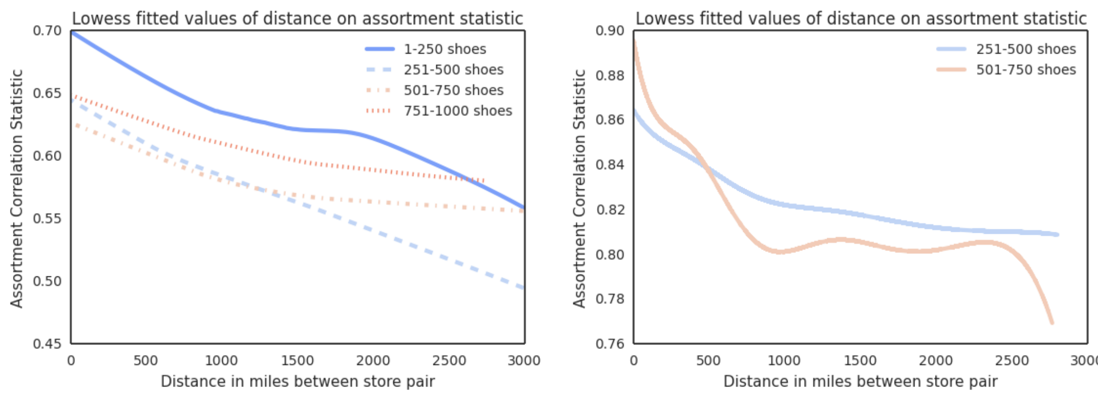
Figure 2: Histogram of percent of stores carrying individual shoes

The plots above suggest heterogeneity in assortments, but do not control for geography. Next, we want to measure how assortments change moving away from a particular store. To calculate this measure, we begin by taking the network of stores and create all possible links. Then for each pair of stores with assortment sets (A, B), we calculate

[10] "We continued to refine and improve the My Macy's process for localizing merchandise assortments by store location, as well as to maximize the effectiveness and efficiency of the extraordinary talent in our My Macy's field and central organization. We have re-doubled the emphasis on precision in merchandise size, fit, fabric weight, style and color preferences by store, market and climate zone. In addition, we are better understanding and serving the specific needs of multicultural consumers who represent an increasingly large proportion of our customers." `https://www.macysinc.com/macys/m.o.m.-strategies/default.aspx`

$$\text{Assortment Overlap} = \frac{\#(A \cap B)}{\min\{\#A, \#B\}}$$

This measure is bounded between zero and one. We use the minimum cardinality, rather than the cardinality of the union in the denominator, because we want this measure to capture differences in the composition of each store's inventory, not differences in assortment size. To further isolate differences in variety from differences in assortment size we directly compare only locations with similar sizes. Figure 3 plots Lowess fitted values for this exercise for Macy's and Payless as a function of distance between stores A and B.



(a) Macy's

(b) Payless

Figure 3: Assortment Overlap by Distance

We see can that the assortment overlap has a decreasing relationship with distance, which suggests these retailers are localizing their product assortments. We also note that as distance approaches zero, assortment similarly does not converge to 1. This is likely reflective of a strategy to increase variety within a geographic area, in addition to locations where retailers created separate men's and women's stores.

2.3 Across-Market Demand Heterogeneity

With our online retail data, prices, product characteristics, and choice sets are the same for all markets. This suggests differences in observed local market shares can only be rationalized by differences in local demand. In Table 2 we present the local and national share of revenue generated by the top 1,000 products ranked within local market. If demand was homogeneous across markets, we would expect the share of revenue accruing to these products to be the same locally and nationally. Thus, the two columns of Table 2 would be equal. Instead we see the share of revenue generated by these products are very large at the local market level compared to their share of revenue at the national level. For example, the top 1,000 products ranked at the metro (combined statistical area - CSA) level make up 86.9% of revenue at the metro level, but these same products only account for 11.5% of national revenue. This suggests that the commonality, even among the most popular products, is quite small across markets.

Table 2: Revenue Share of Top 1,000 Products

Market Definition	Number of Markets	Market Top 1,000	
		Market	National
Combined Statistical Area	165	86.9	11.5
State (plus DC)	51	55.4	19.5
Census Region	4	30.5	24.1
National	1	27.8	27.8

Revenue share of the top products ranked by market and ranked nationally for various levels of geographic aggregation. If demand was homogeneous across markets revenue shares would be equal across columns.

We can formally test for across-market demand heterogeneity using multinomial tests comparing local market shares ($s_{\ell j}$) to national market shares (s_j), where the null hypothesis is $H_0 : s_{\ell j} = s_j$, for all $j \in J$. Table 3 presents the rejection rates for various levels of aggregation. We can see that these tests are overwhelmingly rejected at all levels of aggregation. However, in the tests at the monthly level, we can see the effects of both zeros and aggregation beginning to appear. At more disaggregated levels, zeros become more prevalent, reducing the power of the multinomial tests. On the other end of the spectrum, aggregating up to Census Regions greatly obscures heterogeneity across markets leading to a reduction in rejection rates when compared to the Census Division level.

Table 3: Multinomial Tests - Rejection Rates

	CSA	State	Census Division	Census Region
Month	80.1	89.1	97.6	92.9
Annual	89.3	1	1	1

Rejection rates for multinomial tests comparing local market shares to national market shares. The null hypothesis is $H_0 : s_{\ell j} = s_j$, for all $j \in J$

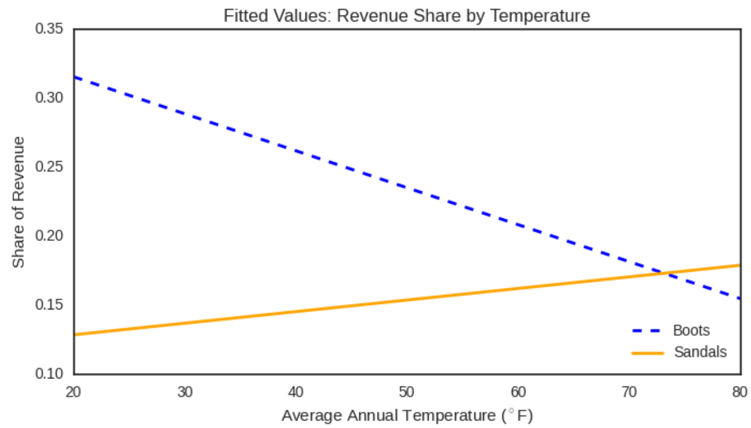
Some differences across markets occur for obvious reasons. Take our earlier example of boots versus sandals. Figure 4 plots the predicted values from a regression of a state's average annual temperature on the share of state revenue captured by boots and sandals. As expected, boots make up a greater share of revenue in colder states and a smaller share in warmer states. Conversely, the opposite relationship holds for sandals.[11]

Other differences across markets occur for less obvious reasons. In Figure 5, we map the consumption pattern of a popular brand by national revenue. Annual sales are mapped into 3 digit zip codes for the eastern United States.[12] While this brand tends to be popular over a large portion of the country, we can see a clear preference for this brand in the northeast. In Florida this brand makes up less than 2.5% of sales, while in parts of New York, New Jersey, and Massachusetts it makes up over 6% of sales. We will exploit this variation to help us identity across-market demand heterogeneity.

[11]This also demonstrates that consumers do not shop online just for products that are not available in traditional brick-and-mortar stores. For example, boots – rather than sandals – make up a sizable share of revenue in Alaska.

[12]We isolate the eastern United States to be able to distinguish differences at fine levels of disaggregation

Figure 4: Boots vs. Sandals Revenue by Temperature



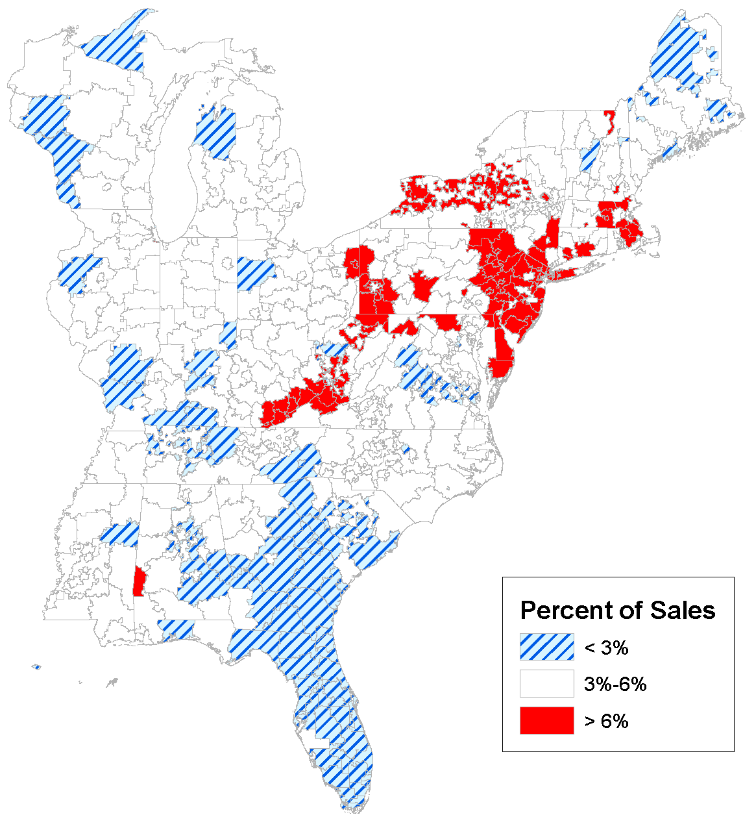
2.4 Aggregation and the Zeros Problem

While demand varies across locations, the data at disaggregated levels exhibits a severe small samples problem, which manifests itself in the form of a zeros problem. Table 4 illustrates the effect of disaggregating the data across both geography and time. For each product, an observation is the number of sales by geographic area and time horizon. We then calculate the percentage of observations where no sale is observed. For example, at the CSA level 95% of products have zero monthly sales. This highlights a small sample problem that is common in high frequency sales data and data with large choice sets.

Observations of zero sales is problematic from both a theoretical and empirical point of view. The distributional assumptions in many demand models imply no products will have zero sales and empirical techniques often require taking the natural log of sales or market shares. This, of course, does not exist when zero sales are observed and common solutions to this problem will lead to biased demand estimates. An in-depth discussion of these issues can be found in Berry, Linton, and Pakes (2004), Gandhi, Lu, and Shi (2013),

and because the interesting portion of the map happens to be the northeastern part of the country

Figure 5: Sales Share of a Popular Brand Across Zip3s



and Gandhi, Lu, and Shi (2014).

Table 4: Data Disaggregation: The Zeros Problem

	Avg. No. of Products	Percent with Zero Sales CSA	State	Region	National
Month	62,768	95.0	85.3	23.3	12.3
Annual	117,493	82.2	59.8	5.7	1.2

Percent of products observed to have zero sales. An observation corresponds to sales at the time(rows)-geography(columns)-product level. CSA: Combined Statistical Area. Region: U.S. Census Region.

On the other hand, aggregation can resolve some of the small sample issue, but it is unsatisfactory because it significantly obscures across-market heterogeneity. For example, we could further aggregate over geography to the Census Region, which would reduce the percentage of zeros to 23.3%. However, this would also reduce the number of markets to four and yet, the percentage of zeros is still quite high, and further aggregation would be necessary. Furthermore, from Table 2, we can see that the top 1,000 products ranked at the Census Region level make up a similar share of revenue at the market and the national levels. That is, we have greatly obscured the across-market heterogeneity of interest to us.

We could also aggregate over product space. Table 5 shows the percentage of zeros and the revenue shares of the top products ranked by market and ranked nationally for products at the SKU-style (our definition of a product) and aggregated to the SKU, brand-category, and brand levels. Since aggregating to the brand-category and brand levels greatly reduces the number of products, we adjust the benchmark to the top 10 "products" rather than the top 1,000.

Table 5: Revenue Share of Top Products: Product Aggregation

Product Definition	Percentage Zeros	Market Top 1,000 Market	Market Top 1,000 National
SKU-style	95	86.9	11.5
SKU	90	92.2	27.4
		Market Top 10 Market	**Market Top 10** National
Brand-Category	77	36.1	24.0
Brand	59	42.8	31.8

Time horizon fixed at monthly level and geography aggregated to the CSA level. Level of aggregation is increasing from top to bottom. This table illustrates how product aggregation lessens burden of small sample sizes but obscures heterogeneity across markets.

The table shows a clear trade-off: At increasing levels of aggregation, the zeros problem is reduced, but this is at the expense of obscuring potential heterogeneity. Similar to aggregation in geography, we see that additional aggregation is still necessary to fully address the zeros problem. However, continued aggregation in either dimension would only further obscure the heterogeneity in which we are interested. This motivates the need for an alternative to aggregation for addressing small sample sizes in our analysis.

3 Model and Estimation

In this section, we present our discrete choice demand model, introduce our proposed estimation technique, and discuss identification. Our modeling approach will closely follow the discrete choice literature. In particular, it will have the form of a nested logit model.[13] We then show how we can use local sales information to identify the across-market demand heterogeneity as a product-market level random effect.

3.1 Model

Each consumer solves a discrete choice utility maximization problem: Consumer i in location ℓ will purchase a product j if and only if the utility derived from product j is greater than the utility derived from any other product, $u_{i\ell j} \geq u_{i\ell j'}, \forall j' \in J \cup \{0\}$, where J denotes the choice set of the consumer and 0 denotes the option of not purchasing a product. For a product j, the utility of a consumer $i \in I$ in location $\ell \in L$ is given by

$$u_{i\ell j} = \delta_j + \nu_{i\ell j}$$

where δ_j is the mean utility of product j for the (national) population of consumers and $\nu_{i\ell j}$ is a random utility component that is heterogeneous across consumers and locations.

[13] A previous version of this paper included a simple logit version of the model. Results are available upon request.

The mean utility of product j is linear in product characteristics and can be written as

$$\delta_j = x_j\beta - \alpha p_j + \xi_j,$$

where x_j is a vector of product characteristics, p_j is the price of product j, and ξ_j is the unobserved (national) product quality for product j. These preferences are constant across locations. The outside good has utility normalized to zero, i.e. $\delta_0 = 0$.

We pursue a nested demand system, where products can be grouped into mutually exclusive and exhaustive sets. Let c denote a nest and note that product j implicitly belongs to a nest c. The outside good belongs to its own nest. We use the category of shoes to define the nesting variable (e.g. boots, sandals, sneakers, etc.). We assume the random utility component can be decomposed into

$$\nu_{i\ell j} = \eta_{\ell j} + \zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}$$

where $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution and, for a consumer i, ζ_{ic} is common to all products in the same category and has a distribution that depends on the nesting parameter λ, $0 \leq \lambda < 1$. $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$ is then also Type-1 extreme value distributed, leading to the frequently used nested logit demand model. λ determines the within category correlation of utilities. When $\lambda \to 1$ consumers will only substitute to products within the same group and when $\lambda = 0$ the model collapses to the simple logit case.

The terms entering $\nu_{i\ell j}$ decompose the heterogeneity in the random utility among consumers into an "across-market" effect, $\eta_{\ell j}$, and a "within-market" effect, $[\zeta_{ic}+(1-\lambda)\varepsilon_{i\ell j}]$. When $\eta_{\ell j} = 0$ for all $\ell \in L$, $j \in J$, then the model reduces to a standard nested logit model, where there is no distinction between local and national preferences.

Integrating over the Type-1 extreme value error terms forms location-specific choice probabilities. These choice probabilities are a function of location-specific mean utilities,

$\delta_{\ell j}$, as well as the substitution parameter λ. The mean utilities are inclusive of the unobservables $(\xi_j, \eta_{\ell j})$. Similar to Berry (1994), the choice probabilities have the following analytic expression:

$$\pi_{\ell j} = \frac{\left(\sum_{j \in c} \exp\{\delta_{\ell j}/(1-\lambda)\}\right)^{1-\lambda}}{1 + \sum_{c' \in C}\left(\sum_{j' \in c'} \exp\{\delta_{\ell j'}/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{\exp\{\delta_{\ell j}/(1-\lambda)\}}{\sum_{j' \in c} \exp\{\delta_{\ell j'}/(1-\lambda)\}} \tag{3.1}$$

$$= \pi_{\ell c} \cdot \pi_{\ell j/c}. \tag{3.2}$$

We then aggregate the location-specific choice probabilities to the national level using the distribution of consumers across locations

$$\pi_j = \int_L \pi_{\ell j} dF\omega = \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell j},$$

where $dF\omega$ is the density of location population shares and, in discrete notation, ω_ℓ is the population share of location ℓ.

As shown in Berry (1994) and Berry, Levinsohn, and Pakes (1995), we can invert Equation 3.1 and proceed with linear instrumental variables methods to estimate the preference parameters as well as estimate ξ_j as fixed effects. The local level residuals would form estimates of $\eta_{\ell j}$. However, the sparsity of sales of individual products across locations would lead to attenuation bias. To circumvent this problem, we use a random-effects specification, where $\eta_{\ell j}$ is drawn independently from a normal distribution, $N(0, \sigma_j^2)$. Instead of attempting to recover $\eta_{\ell j}$ from the data, we instead estimate σ_j^2. We maintain the fixed effects assumption for ξ_j. Since our online retailer sets national prices, prices may be correlated with xi_j, but we assume characteristics are exogenous with respect to η. Finally, we assume that both π_j and $\pi_{\ell c}$ are well approximated given our sales data, but sales of products at specific locations do not well approximate $\pi_{\ell j}$.

3.2 Inverting the Market Share

We now show that the inverse of our market shares takes a convenient analytical form, which will simplify the simulation of our local choice probabilities. While small sample sizes make local observed market shares for individual products unreliable estimates of the underlying choice probabilities, we assume the local choice probabilities of the outside good, $\pi_{\ell 0}$, and the category shares, $\pi_{\ell c}$, are well estimated by the data.[14] We present our market share inversion in the following proposition:

Proposition 1. *For any set of $\{\eta_\ell\}_{\ell=1}^{L}$ the market share inversion takes the following analytic form, $\forall j \in J$,*

$$\delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right). \tag{3.3}$$

Proof. In the nested logit case, we will find it convenient to write shares as a fraction of the category share. By Bayes' rule

$$\begin{aligned}\pi_j(\eta_\ell;\delta,\lambda) &= Pr_\ell\{\text{ c }\}\cdot Pr_\ell\{\text{ j} \mid \text{c }\}\\ &= \pi_{\ell c}\cdot\frac{\exp\{(\delta_j+\eta_{\ell j})/(1-\lambda)\}}{\sum_{j'\in c}\exp\{(\delta_{j'}+\eta_{\ell j'})/(1-\lambda)\}},\end{aligned}$$

Aggregated choice probabilities are then

$$\pi_j = \sum_{\ell=1}^{L}\omega_\ell\pi_j(\eta_\ell;\delta,\lambda) = \sum_{\ell=1}^{L}\omega_\ell\pi_{\ell c}\frac{\exp\{(\delta_j+\eta_{\ell j})/(1-\lambda)\}}{\sum_{j'\in c}\exp\{(\delta_{j'}+\eta_{\ell j'})/(1-\lambda)\}}.$$

Next, define

$$D_{\ell c} = \sum_{j'\in c}\exp\left\{\frac{\delta_j+\eta_{\ell j}}{1-\lambda}\right\}.$$

We normalize the utility of the outside good – both in terms of product characteristics, as well as, the unobserved taste preference across locations. This means the probability of choosing the outside good at location ℓ is equal to

$$\pi_{\ell 0} = \frac{1}{1+\sum_{c'\in C}D_{\ell c'}^{1-\lambda}},$$

[14]The populations of CSAs are fairly large, so we believe the law of large numbers applies for the decision to purchase versus not to purchase and for the decision among a small number of categories. However, the number of purchases compared to the number of products is small, so we cannot apply the law of large number to the sales of individual products.

and note that the probability of choosing a good in category c at location ℓ is equal to

$$\pi_{\ell c} = \frac{D_{\ell c}^{1-\lambda}}{1 + \sum_{c' \in C} D_{\ell c'}^{1-\lambda}},$$

thus

$$D_{\ell c} = \left(\frac{\pi_{\ell c}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda}}.$$

Plugging into the aggregate choice probabilities, gives

$$\begin{aligned}
\pi_j &= \sum_{\ell \in L} \omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\delta_j + \eta_{\ell j}}{1-\lambda}\right\} \\
&= \exp\left\{\frac{\delta_j}{1-\lambda}\right\} \sum_{\ell \in L} \omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}
\end{aligned}$$

Finally, taking logs we have our inversion:

$$\log\left(\pi_j\right) = \frac{\delta_j}{1-\lambda} + \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)$$

or

$$\delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right)$$

■

Since the population shares, ω_ℓ, and the shares $\pi_{\ell 0}$ and $\pi_{\ell c}$ are known, this equation relates δ_j to the aggregated share data, π_j. Additionally, notice that this reduces to the standard Berry (1994) inversion when $\eta_{\ell j} = 0, \forall \ell, j$. In the next subsection, we describe how we use our inversion estimate the demand parameters. To identify the distribution of across-market heterogeneity, we will introduce a set of micro-moments created from information on local level sales. We can then integrate out this distribution, according to our inversion, to obtain the mean utilities, δ_j, from the data, π_j, and proceed with traditional estimation techniques at the aggregate level.

3.3 Estimation

Suppose we knew, or had an estimate for, $\sigma = \{\sigma_j\}_{j=1}^{J}$. Then by simulating $\tilde{\eta}_{\ell j} \sim N(0, \sigma_j^2)$, we can exploit the structure of the model. By law of large numbers,

$$\pi_j \approx \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell j}(\tilde{\eta}_\ell; \delta, \lambda),$$

so long as the number of locations L is sufficiently large. Thus, aggregated choice probabilities only depend on the variance of the across-market heterogeneity, σ, rather than on than the individual fixed effects, η_ℓ, themselves. Therefore, national demand can be expressed as

$$\pi_j = \pi_j(\delta, \lambda; \sigma),\ \ j = 1, ..., J,$$

which is a system of equations that can, in general, be inverted (Berry, Gandhi, and Haile 2013) to yield,

$$\delta(\pi, \lambda, \sigma) = x_j\beta - \alpha p_j + \xi_j.$$

Following BLP, for a fixed σ, we can use linear instrumental variables z_j, such that $E[z_j\xi_j] = 0$ and $E[z_j'(p_j, x_j)]$ has full rank, to identify (α, β) as a function of σ. However, the existing instruments used in the literature[15] typically provide little to no identifying power for the non-linear parameter σ (Gandhi and Houde 2014). Instead we use the disaggregated information in our data to augment the instrumental variable conditions with an additional set of micro-moments that provide direct information on σ (Petrin 2002).

[15] For example, BLP instruments

3.4 Micro Moments

Let $P0_{\ell j}(\sigma;\delta,\lambda)$ be the probability that a product j has zero sales given the N_ℓ consumers observed to purchase a shoe in location ℓ. We then define,

$$P0_j(\sigma;\delta,\lambda) = \frac{1}{L}\sum_{\ell=1}^{L} P0_{\ell j}(\sigma;\delta,\lambda)$$

to be the fraction, or proportion, of markets that the model predicts will have zero sales for product j. Observe that this fraction depends on model parameters, where we have concentrated out δ as $\delta(\pi,\lambda,\sigma)$. The empirical analogue is

$$\hat{P0}_j = \frac{1}{L}\sum_{\ell=1}^{L} 1\{s_{\ell j} = 0\},$$

where $s_{\ell j}$ is the observed location level market share for product j. Our micro-moment then identifies σ by matching the model's prediction to the empirical analogue, i.e.

$$mm(\sigma;\delta,\lambda) = \frac{1}{J}\sum_{j=1}^{J}\left(P0_j(\sigma;\delta,\lambda) - \hat{P0}_j\right).$$

It is important to point out that $P0$ is just one such micromoment that can be used to estimate across market demand heterogeneity. Other moments include $P1, P2$, etc., as well as variance in sales across markets. Note that $P0$ remains valid as the number of locations increases. This is because we assume finite population for a given market which implies as $L \to \infty$, a positive proportion of locations may experience zero sales for a given product.[16]

[16] In Monte Carlo studies, we have found adding additional micro-moments does not greatly affect the estimates. Also, the logit structure implies $P0$ is no longer valid when assuming large N for all locations since then each product will have positive share. We estimate these models as well.

3.5 Estimation Procedure

Having laid the foundation of our methodology, we turn to detailing the computational mechanics of the estimation. The model can be estimated using two-step feasible generalized method of moments (GMM). We start with the implementation of our micro-moments. Note that local level mean utilities can then be written as

$$\delta_{\ell j} = \delta_j + \eta_{\ell j} = \delta_j + \sigma_j \bar{\eta}_{\ell j}$$

where $\bar{\eta}_{\ell j}$ is an i.i.d. draw from a standard normal distribution. The nested logit structure implies

$$\hat{\pi}_{\ell j} = \pi_{\ell c} \frac{\exp\{(\delta_j + \sigma_j \bar{\eta}_{\ell j})/(1-\lambda)\}}{\sum_{j' \in c} \exp\{(\delta_{j'} + \sigma_{j'} \bar{\eta}_{\ell j'})/(1-\lambda)\}}. \tag{3.4}$$

The local level choice probabilities can be used to simulate consumer purchases at each location, holding the number of observed purchases, N_ℓ, fixed. In particular, the probability a product is observed to have zero sales at location ℓ is

$$P0_{\ell j}(\sigma; \delta, \lambda) = (1 - \hat{\pi}_{\ell j})^{N_\ell},$$

i.e. the probability we observe N_ℓ sales, none of which are good j. Averaging over locations allows us to match the proportion of locations observing zero sales of j. These are our micro-moments, $mm(\cdot)$. This is computationally fast and avoids the problems posed by simulating individual purchase decisions. Our micro-moments allow us to pin down the across market heterogeneity parameters (σ) while explicitly accounting for small samples at the location level – this is coming from modeling the N_ℓ purchase decisions by location.

With a candidate solution of σ, the structure we have placed on the η's allows us to integrate them out by subtracting the sum of local random effects according to the market share inversion in Equation 3.3.[17] However, in order to recover δ, we also need to have

[17]Since we take many draws over the distribution of $\eta_{\ell j}$, Proposition 1 implies that we can estimate the sum

a candidate λ, suggesting the estimation procedure must also search over the nesting paramter.

Specifying a σ and λ allows us to recover

$$\delta_j = x_j\beta - \alpha p_j + \xi_j,$$

where we can then use standard instrumental variables methods to control for price endogeneity. Let Z be the matrix of instruments to identify β, α. Another set of moments to include are $m_1 = \mathbb{E}[Z'\xi]$.

The last complication to address is how to identify the nesting parameter. In the Berry (1994) nested logit inversion, within category shares are also correlated with the unobserved product quality creating an endogeneity problem. A similar issue arises in our inversion. Note that, with δ as defined in Equation 3.3,

$$E\left[\frac{\partial\delta_j(\pi,\lambda,\sigma)}{\partial\lambda}\cdot\xi_j\right] \neq 0$$

because ξ_j enters the aggregate product share, π_j, and the local level category shares, $\pi_{\ell c}$. Berry (1994) solves this problem by employing an instrument, $z_{j|c}$, that is correlated with the within category share, but uncorrelated with the unobserved product quality.[18] The same instrument can be employed here, since $z_{j|c}$ is correlated $\frac{\partial\delta_j(\pi,\lambda,\sigma)}{\partial\lambda}$ through the local level category shares, but still uncorrelated with the unobserved product quality. Thus, if $z_{j|c}$ is a valid and relevant instrument when estimating the nested logit model using the Berry (1994) inversion, it is a valid and relevant instrument for our inversion. These moments are m_2.

in Equation 3.3 without explicitly knowing each individual $\eta_{\ell j}$. Here we are assuming LLN in L.

[18] For example, a combination of the product characteristics of competing products within the same category or nest.

Stacking moments where $\theta = \sigma, \lambda, \beta, \alpha$ we have

$$G(\theta;\cdot) = \begin{bmatrix} mm \\ m_1 \\ m_2 \end{bmatrix}$$

and the GMM criterion is $G(\theta;\cdot)^T W G(\theta;\cdot)$, with weighting matrix W. We first take $W = I$ and then use $W = \left(G(\theta;\cdot)G(\theta;\cdot)^T\right)^{-1}$ in the second step.

Included in x is product ratings for comfort, look, and overall appeal, and fixed effects for color, brand, and time. We instrument for price using the characteristics of competing products (BLP instruments), grouped by brand. That is, let B denote the set of brands and let J_b denote the set of products manufactured by brand $b \in B$, then, for each time period, our set of instruments is

$$x_j, \quad \sum_{j' \neq j}^{J_b} x_j, \quad \sum_{j'=1}^{J_{-b}} x_j.$$

Additionally, to identify the nesting parameter, λ, we use the characteristics of competing shoes by category. Let c_b denote the set of shoes manufactured by brand $b \in B$ in category $c \in C$, then, for each time period, we include the following additional instruments

$$\sum_{j' \neq j \in c_b} x_{j'}, \quad \sum_{j' \in c_{-b}} x_{j'}.$$

Finally, we parametrize σ as[19]

$$\sigma_j = h(\text{category}_j) = \gamma_c.$$

[19]The heteroskedasticity function cannot depend on location-specific information since the estimation procedure applies a law of large numbers on L. In addition to category, we have estimated the model using a parametric function of product rank, as well as interacting rank and category fixed effects. The results are similar to what we present here. We have noted more complicated functions, such as polynomials of rank interacted with category information are too computationally burdensome.

3.6 Identification

The variance of our local level random effect, $\eta_{\ell j}$, is identified through differences in local market shares. Ideally, if we had a large sample of sales and there were no across-market demand heterogeneity, each product's local market shares would be the same in every market, and our variance would be zero. However, differences in local market shares may arise due to sampling, particularly in the presence of small sample sizes. Therefore, in our construction of the micro-moments, we are careful to account for the number of sales in each market.

For each product, we will use the proportion of locations in which zero sales for a product are observed to form our micro-moments. To understand the intuition behind this, consider a world with a single inside good. If demand is homogeneous across markets, at the disaggregated level, we would expect to see similar local shares.[20] In particular, if this good is very popular at the aggregate level, we would expect to observe few, if any, local markets with zero sales. Instead suppose we observe wildly different shares across markets with a significant portion of markets having zero sales. This suggests the product faces heterogeneous demand across markets. Assuming a normal distribution, as we do, the variance of this heterogeneity can then be pinned down by the number of observed zeros. If a large number of zeros are observed, this suggests a large number of markets drew low valuations for the good (a low draw of $\eta_{\ell j}$), which suggests a higher variance for the heterogeneity. This is because the higher the variance the greater the density of low $\eta_{\ell j}$ draws. Conversely, few observed zeros suggests there are few markets with low draws of $\eta_{\ell j}$ and, hence, a lower variance.

Parameters within δ_j are identified through the standard channels – in the cross-section through variation in aggregate sales given characteristics, (x_j, p_j), and across time periods through time varying characteristics and variation in the choice set J. Identification of

[20]This is not necessarily the case given the small samples problem, but the micro-moments address this issue.

the nesting parameter, λ, is driven by changes to the category shares as the utilities of products within the category change or the set of products within the category changes.

4 Results

In this section, we discuss our estimates and the fit of the model. We will define our geographic locations to be composed of 165 Combined Statistical Areas (CSAs) and our time horizons to be at the monthly level. We will begin by discussing the demand parameters that are constant across locations. This will allow us to more easily compare estimation results across methodologies and specifications. Then we present our heterogeneity results. We find that accounting for across-market heterogeneity is particularly important for explaining the observed distribution of sales at the local level. In the next section, we will conduct our counterfactual exercises.

4.1 Demand Parameters Constant Across Markets

A summary of our demand estimates is presented in Tables 6 and 7 for men's and women's shoes, respectively. Each specification includes fixed effects for brand, color, and time. We present three sets of estimates: (1) the nested logit demand model estimated at the CSA level, which we will call "local FE" because it produces a location-product level fixed effect; (2) the nested logit demand model estimated at the national level, which we will call "national FE" because it produces an aggregate national-product level fixed effect; and (3) our estimation procedure, which we will call "local RE" because it produces a location-product level random effect. We also account for any remaining zeros using the correction proposed by Gandhi, Lu, and Shi (2014). A discussion of the correction procedure can be found in Appendix A.

Specification (1), the nested logit demand model estimated at the local CSA level, illustrates the selection bias generated by the severity of the zeros problem. When estimating the model at the CSA level, each observation is a product-location specific share. Thus,

Table 6: Demand Estimates - Men's

	Local FE (1)	National FE (2)	Local RE (3)
Price	-0.029	-0.019	-0.024
	(0.001)	(0.000)	(0.002)
Comfort	0.004	-0.018	-0.012
	(0.005)	(0.011)	(0.015)
Look	-0.136	-0.062	-0.080
	(0.005)	(0.012)	(0.018)
Overall	0.209	0.249	0.312
	(0.006)	(0.012)	(0.031)
No Reviews	0.356	0.713	0.936
	(0.018)	(0.044)	(0.091)
Constant	-10.924	-10.091	-11.419
	(0.081)	(0.142)	(0.299)
λ	0.129	0.602	0.512
	(0.004)	(0.011)	(0.006)
σ	—	—	*
Fixed Effects			
Brand	✓	✓	✓
Color	✓	✓	✓
Month	✓	✓	✓
N	28,916,910	175,254	175,254
Zeroes	27,597,784	30,282	30,282
	(95.4%)	(17.3%)	(17.3%)
Price Elast.			
Mean	-3.054	-5.501	-4.833
Std. Dev.	(1.745)	(4.082)	(4.033)

Notes: Estimated at the monthly level. "Local FE" (1) estimates the and nested logit models at the CSA local market level, creating location-product level fixed effects. "National FE" (2) estimates the nested logit model at the national level, creating national-product level fixed effects. Finally, "Local RE" (3) estimates the nested logit model using our estimation technique to allow for across-market heterogeneity in the form of a location-product level random effect.
Standard errors in parentheses. All reported coefficients are significant at the 1% level.
$*$ estimates for across-market heterogeneity in specification (3) will be discussed in the following subsection.

Table 7: Demand Estimates - Women's

	Local FE (1)	National FE (2)	Local RE (3)
Price	0.001	-0.003	-0.005
	(0.000)	(0.000)	(0.000)
Comfort	0.108	-0.021	0.008
	(0.003)	(0.004)	(0.001)
Look	-0.054	-0.014	-0.044
	(0.003)	(0.004)	(0.013)
Overall	0.094	0.025	0.073
	(0.003)	(0.006)	(0.004)
No Reviews	0.147	0.035	0.076
	(0.009)	(0.014)	(0.001)
Constant	-14.629	-17.938	-11.614
	(0.014)	(0.052)	(0.097)
λ	-0.026	0.863	0.648
	(0.001)	(0.008)	(0.024)
σ	—	—	*
Fixed Effects			
Brand	✓	✓	✓
Color	✓	✓	✓
Month	✓	✓	✓
N	57,682,020	349,588	349,588
Zeroes	55,141,243	66,827	66,827
	(95.6%)	(19.1%)	(19.1%)
Price Elast.			
Mean	0.108	-2.866	-1.525
Std. Dev.	(0.071)	(2.456)	(1.339)

Notes: Estimated at the monthly level. "Local FE" (1) estimates the and nested logit models at the CSA local market level, creating location-product level fixed effects. "National FE" (2) estimates the nested logit model at the national level, creating national-product level fixed effects. Finally, "Local RE" (3) estimates the nested logit model using our estimation technique to allow for across-market heterogeneity in the form of a location-product level random effect.
Standard errors in parentheses. All reported coefficients are significant at the 1% level.
$*$ estimates for across-market heterogeneity in specification (3) will be discussed in the following subsection.

the number of observations is 165 times greater (number of products times 165 CSAs) than the other specifications. Unfortunately, at this level of disaggregation about 95% of the observations have zero sales resulting in coefficients that are severely biased toward zero. Of particular concern for us are the price coefficients, which are positive for women, and the nesting parameters, which are severely attenuated for men and negative for women. In the bottom panels of each table, we can see that the local specifications imply price elasticities that are more inelastic than the other specifications. With the result for women being nonsensically positive, implying an increase in price would increase sales.

Specifications (2) and (3) directly compare the results estimated using the standard approach on national level data and the results estimated using our procedure for the nested logit model. Unsurprisingly, the results for these specifications are similar. However, the advantage of our approach is that it retains information on the distribution of heterogeneity across locations. The importance of this distinction will be highlighted in the following section when we perform counterfactual analyses.

Using our estimation technique (Local RE), the price coefficients have the expected signs, -0.024 and -0.006 for men's and women's shoes, respectively. These results suggest that men are far more price sensitive than women when it comes to their footwear purchases with mean product level price elasticities of -4.833 versus -1.525.

Turning to the coefficients on our review variables, we can see that the overall rating has the expected sign in all specifications, with higher ratings having positive effects on demand. Look and comfort, however, appear to have a sign that is opposite from the one expected in at least some specifications. It is possible that, after controlling for the overall rating, the qualities that make a shoe more aesthetically pleasing or more comfortable reduces a shoe's appeal through some other channel.[21] Meanwhile, our indicator for no reviews takes on positive signs for both men's and women's shoes. This variable largely captures the demand for new products before there has been an opportunity to review

[21]It is also likely that after controlling for overall appeal, the estimates for look and comfort suffer from multicollinearity.

them. New products often benefit from additional promotion and advertising and it is likely that the positive effect of having no review actually reflects the additional promotion, rather than a desire to purchase shoes that have not been reviewed.

4.2 Across-Market Heterogeneity

Our results in the previous subsection depend on our estimates of $\sigma_j = h(\cdot)$. We identify the distribution of across-market heterogeneity

$$\sigma_j = h(\text{category}_j) = \gamma_c,$$

by matching the probability, implied by the model, that a product has zero sales with the observed percentage of locations that have zero sales. Our estimates for the across-market demand heterogeneity are presented in Table 8. For comparison, the interquartile range and the standard deviation of the national mean utilities, δ_j, are also reported.

In general, women's shoes tend to exhibit greater across-market demand heterogeneity than men's shoes. This is due in part to the greater number of women's shoe varieties. The categories with the largest across-market heterogeneity conditional on category choice are oxfords (0.822), slippers (0.952), and sneakers (0.815) for men and slippers (1.285) for women. These appear to be quite large with across-market variation in some categories approaching about half the magnitude of the aggregate mean utilities.

Figure 6 illustrates the fit of our estimation technique and gives us further insight into our heterogeneity results. It plots the percentage of location level zero market shares by product. The left panels are plots for men's shoes and the right panels are for women's shoes. For comparison, we include simulation results for the case of homogeneous demand across markets, i.e. when $\sigma_j = 0$ for all shoes j. At the head of the distribution there are fewer location level zero market shares, but, because mean utilities are relatively high, variation is required to produce these zeros. Moving toward the middle of the distribution, this variation is able to account for the increasing percentage of zero market shares. If

Table 8: Results: Across-Market Heterogeneity: $\sigma_j = h(\cdot)$

	Men	Women
Boat	0.652	0.879
Boots	0.248	0.798
Clogs	0.246	0.829
Flats	—	0.582
Heels	—	0.670
Oxfords	0.822	0.262
Sandals	0.568	0.190
Slippers	0.952	1.285
Sneakers	0.815	0.213
δ		
IQR	1.720	2.320
St. Dev.	1.628	1.901

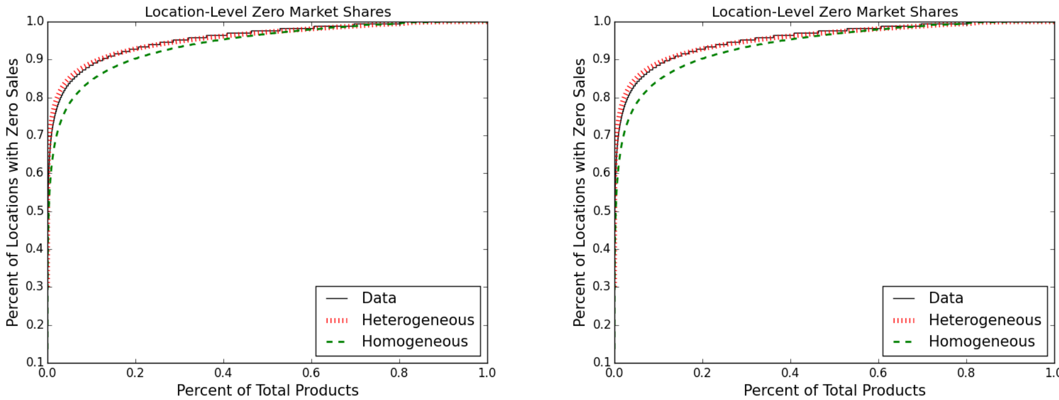
Distribution of across-market heterogeneity. The bottom panel presents summary information on δ for comparisons of magnitudes.

demand were homogeneous across markets, we would expect to see far fewer zeros among popular and mid-ranked products.

5 Analysis of the Estimated Model

In this section we use the estimated model to perform counterfactual analysis under a series of restricted choice sets. We will begin by presenting our primary results, allowing for tastes to differ across markets and for local brick-and-mortar retailers to customize their assortments to local demand (Section 5.1). We will then present results shutting down local assortment customization and show how these results overestimate the gains to online variety (Section 5.2). Finally, we revisit the phenomenon of the long tail and show that aggregation of sales over markets with different tastes is a key driver of the long tail of online retail (Section 5.3).

Figure 6: Goodness of Fit: Percentage of Location Level Zeros



Notes: (left) Men's (right) Women's. For each product, percentage of locations with zero sales in the data (solid), in our estimation with across-market heterogeneity (vertical-dash), and with homogeneous demand across markets (dash).

Since local brick-and-mortar product assortments are often not directly observed by researchers, they must be inferred from the estimated demand system. Consistent with the literature, we assume local brick-and-mortar retailers stock the top K_ℓ most popular products. The ordering of products is determined by the estimated local mean utilities from the demand system. The literature often establishes the same threshold for all markets, but we have more information we can bring to bear. While we cannot directly match our online sales data and our brick-and-mortar assortment data, we can use the counts as a guide to our selection of local level assortment sizes. We also examine the robustness of our results for a range of thresholds in the next section.

Mechanically, to compute our counterfactuals, we draw a set of η_ℓ's for each location. Products are then ranked in each location by their location specific mean utilities and the top products are included in the counterfactual choice set. For each counterfactual choice set, location level choice probabilities are then calculated according to Equation 3.4. Using these probabilities, we simulate location level purchases, which then allows us to compute counterfactual consumer welfare and retail revenue.

5.1 Counterfactuals with Across-Market Heterogeniety

We begin our analysis by performing the counterfactuals for our primary result. In each counterfactual, we restrict the size of the choice set in each market, but each market is allowed to carry the top products specific to that location. Consumer purchasing decisions are then simulated under the restricted choice sets. For each counterfactual scenario and specification, we calculate: location level consumer welfare

$$CS_\ell = \frac{M\omega_\ell}{\alpha} \log\left(1 + \sum_{c \in C} \left(\sum_{j \in c} \exp\{\delta_j + \eta_{\ell j}\}\right)^{1-\lambda}\right)$$

and retail revenue,

$$r_{\ell j} = p_j M \omega_\ell \pi_{\ell j},$$

where M is the size of the national population. For each of our specifications, Table 9 presents the increase in consumer welfare from online variety and Table 10 presents the consumer welfare and retail revenue under the restricted choice set relative to the unconstrained online choice set.

Examining the results of our estimation technique, we find that when consumers are able to move from a world where they only have access to the goods available at their local stores to a world where they have access to the whole online choice set, consumer welfare increases by 6.8% or $57.20 million.

The deficiencies of using the standard approach with location level market shares is also highlighted by Table 9. Consistent with Ellison and Glaeser (1997), ignoring the local level small samples problem exaggerates the extent of heterogeneity across markets. By assuming products without an observed sale are completely unwanted at that particular location, it is "easier" for our hypothetical brick-and-mortar retailers to customize their

Table 9: Local Choice Set: Consumer Welfare Increase

	Local FE	Local RE
Increase		
- Absolute ($ Millions)	0.72	57.20
- Percent	~ 0%	6.8%

assortments to local demand. This leads us to understate the consumer welfare increase. In our application, using the standard approach with location level market shares implies that the consumer welfare gains from online variety are almost nonexistent. This is because, in our data, relatively few shoe varieties makeup the observed sales in individual local markets. Thus relatively few shoe varieties are required to completely satisfy local demand when ignoring the small samples problem. Note that we omit the national level specifications. While these specifications may be consistent with across-market demand heterogeneity, there is no way to determine the underlying geographic distribution of this heterogeneity.

Table 10: Local Choice Set: Share of Unconstrained

	Local FE	Local RE
Consumer Welfare	~ 100%	93.6%
Revenue	~ 100%	92.2%

Table 10 suggests that consumer welfare derived from access to online variety may be surprisingly small. Using our preferred specification, local RE, if local stores stock products that target local demand, consumers would capture 93.6% of the unconstrained consumer welfare, the total consumer welfare that would be obtained with access to all of the products. Conversely, having access to the entire online choice set only accounts for 6.8% of the total unconstrained consumer welfare. Similar conclusions can be drawn for retailer revenue. A national brick-and-mortar chain can generate 92.2% of the total revenue it would generate from stocking the universe of products, by stocking a only small number of well selected products. Also, Table 10 again highlights the effects of ignoring the small samples problem. Similar to the result for consumers, estimates ignoring the small samples problem imply almost zero benefits to firms of stocking the entire universe of varieties.

5.2 Counterfactuals with Nationally Standardized Choice Sets

In this subsection, we perform counterfactual analyses similar to the ones above. However, we impose the additional constraint that each market will be restricted to the top products determined by ranking products according to their national mean utilities, δ_j. Using our preferred estimates, Table 11 presents the increase in consumer welfare from online variety and Table 12 presents the consumer welfare and retail revenue under the restricted choice set relative to the unconstrained online choice set.

Table 11: National Choice Set: Consumer Welfare Increase (Local RE)

	Local Choice Set	National Choice Set
Increase		
- Absolute ($ Millions)	57.20	146.63
- Percent	6.8%	19.5%

Table 11 shows that failing to account for customization in local assortments will overstate the gains to consumer welfare. Under a standardized national assortment, access to online variety increases consumer welfare by 19.5% or $146.63 million. This suggests failing to account for heterogeneity across markets will overstate consumer welfare due to online variety by 187% in percentage terms and 156% in absolute terms. The overstatement occurs because the initial welfare (pre-internet) of consumers is lower when choice sets are nationally standardized than when they are locally targeted, which can be seen in Table 12.

Note that our results are nearly identical, whether the model is estimated using standard approaches with aggregate (national) level market shares or our estimation method, so we omit them here. This is unsurprising given our demand results and because, under both specifications, consumers from different locations are pooled into a single population at the national level. The drawback of using national level market shares, however, as previously mentioned, is that there is no way to determine the underlying geographic distribution of heterogeneity preventing local level analysis.

Table 12: National Choice Set: Share of Unconstrained (Local RE)

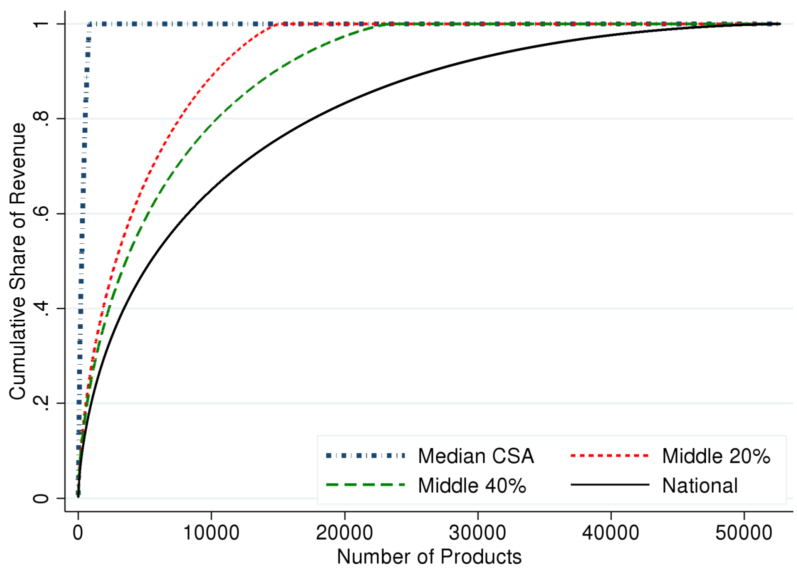
	Local Choice Set	National Choice Set
Consumer Welfare	93.6%	83.7%
Revenue	92.2%	80.5%

Turning to retailer revenue we see that a researcher assuming a nationally standardized assortment will severely underestimate counterfactual brick-and-mortar revenue. A national brick-and-mortar chain would generate just 80.5% of total unconstrained revenues with a nationally standardized assortment compared to 92.2% of total unconstrained revenues with locally customized assortments. This suggests that there is a significant incentive for local stores to cater to local demand. By doing so they would obtain 14.5% greater revenue than under a nationally standardized assortment.

5.3 Long Tail Analysis

Our counterfactual results in the previous two subsections suggest that "shorter" tails at the local level underly the long tail at the national level. Using the raw sales data, Figure 7 illustrates how local level "short" tails can aggregate to a national level long tail. It plots the cumulative share of revenue going to the top K products for the median CSA (by number of monthly sales), middle 20%, middle 40%, and national level markets. For a single local market, we can see that there is an extremely short tail with fewer than 3,000 products making up all the of sales in that CSA. Since the popularity of products varies wildly across geographic markets, aggregating over markets increases the number of different varieties sold and decreases the density of sales among the top ranked products. Sales becoming less concentrated among the top products produces a lengthening effect on the tail of the sales distribution.

Figure 7: Aggregating to the Long Tail



However, the small samples problem in the raw data presents us with a skewed perspective in that it suggests a ridiculously short tail at the local level. Using our estimated model, we can correct for the small samples problem in our long tail analysis by simulating a large number of sales in these markets. Figure 8 plots the cumulative share of revenue going to the top K products for the median CSA correcting for the small samples problem.

As expected, we find that the local tail is actually quite a bit longer than suggested by the raw data.

Figure 8: Local Tail: Correcting for Small Samples

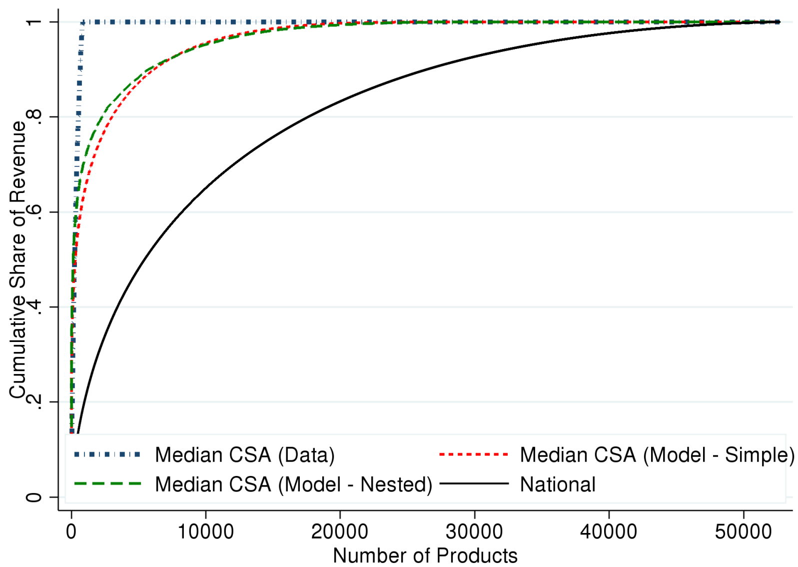


Table 13 further illustrates the effect of small samples on the local tail. It presents the average share of revenue accruing to products outside of the top 3,000 products. At the national level, more than 50% of revenue comes from products ranked outside the top 3,000. At the local level, if we were to rely on the raw data, we would find that only 3.4% of revenue comes from products ranked outside the top 3,000 products. In other words, 96.6% of demand could be satisfied with just 3,000 well targeted products in each market. This may lead us to erroneously conclude that there is no long tail as described in the existing literature. Simulating our model with the same small number of sales in each local market yields very similar results.[22] However, by simulating a large number of sales

[22]Given the small number of sales at the local level, this result is unsurprising. For example, suppose fewer than 3,000 sales are observed in a local market. Then, of course, the share of revenue going to products outside the top 3,000 is zero.

in each local market, we find that there is, in fact, significant demand for niche products at the local level with 23.6% of sales coming from products outside the top 3,000 products.

Table 13: Average Revenue Share of Products Outside of the Top 3,000

	Small Sample		Large Sample
	Data	Model	Model
National	50.6%	49.4%	50.6%
Local	3.4%	2.6%	23.6%

6 Robustness

In this section, we examine the robustness of our findings to the size of the counterfactual choice set. While we find that the size of the overstatement is sensitive to the size of the counterfactual assortment size, our findings from the previous section are on the lower end, suggesting our conclusions are on the conservative side. Table 14 presents the change in consumer welfare and the size of the overstatement resulting from various thresholds of the counterfactual choice set, respectively. For comparison, we also include our baseline results from the previous section.

Unsurprisingly, as the size of the counterfactual choice set increases the gain to consumers from access to the remaining products decreases. This decrease occurs substantially faster under locally customized assortments than under a nationally standardized assortment. As a result, the percentage overstatement is increasing in the assortment size, despite the absolute size of the overstatement decreasing. This pattern is illustrated in Figure 6. Figure 6 can be read as the estimated consumer welfare overstatement when assuming no local assortment customization, measured in millions of dollars (solid) and as a percentage (dash). The absolute overstatement peaks at $207.9 million with about about 2% of products (or 1,000 shoe varieties). The percentage overstatement approaches

Table 14: Robustness: Overstatement of Consumer Welfare Increase

Assortment Size	Percent Increase Loc.	Nat.	%Δ	Absolute ($ Millions) Loc.	Nat.	Δ	%Δ
Baseline	6.8	19.5	186.8	57.20	146.63	89.43	156.35
Threshold							
3,000	24.6	71.1	189.0	354.19	372.74	18.55	5.24
6,000	13.7	45.6	232.8	216.61	281.11	64.50	29.78
12,000	4.7	19.8	321.3	81.00	148.13	67.13	82.88
24,000	0.8	5.6	600.0	15.01	47.90	32.89	219.12

infinity as the fraction of available products approaches one.

Figure 9: Overestimation of Consumer Welfare

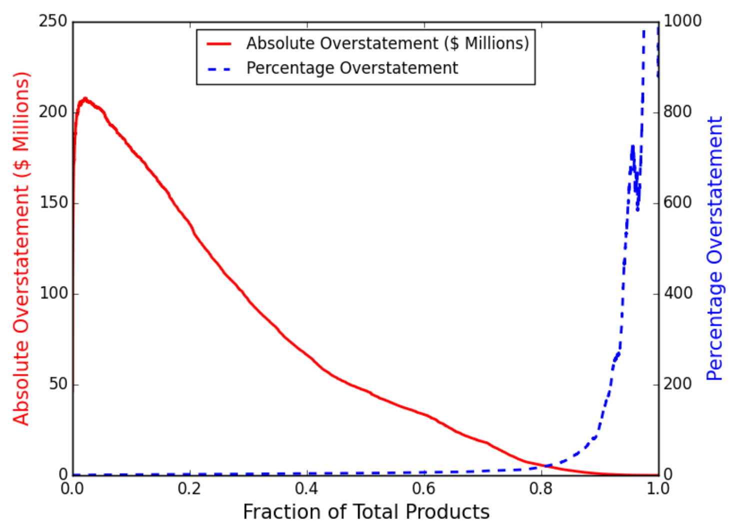


Table 15 presents the retail revenue at various thresholds of the counterfactual choice set. With retail revenue we find that as assortment sizes increases the gain from customizing assortments to local demand is decreasing. However, a typical large brick-and-mortar shoe retailer stocks, at most, a few thousand varieties. This puts them at the small end of our robustness analysis, suggesting there may be significant incentives for large national brick-and-mortar shoe retailers to customize their assortments to local demand.

Table 15: Robustness: Retail Revenue

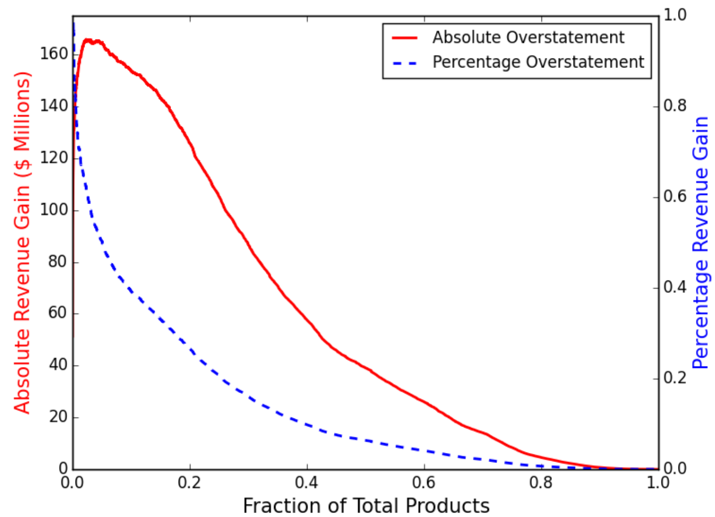
	Absolute ($ Millions)			
Assortment Size	Loc.	Nat.	Δ	$\%\Delta$
Baseline	610.68	533.57	77.11	14.45
Threshold				
3,000	506.34	344.55	161.79	46.96
6,000	559.42	408.78	150.64	36.85
12,000	620.94	523.99	96.95	18.50
24,000	655.43	622.20	33.23	5.34

Figure 6 graphs the increase in retail revenue due to local customization of assortments, measured in millions of dollars (solid) and as a percentage (dash). The absolute gain in revenue from localization peaks at $165.9 million at about 3% of products (or 1,500 shoe varieties). The percentage gain is monotonically decreasing with assortment size. The graph shows that when assortment sizes are extremely limited, brick-and-mortar retailers can significantly boost revenue by maintaining locally customized product assortments.

7 Conclusion

In this paper, we quantify the effect of increased access to variety due to online retail on consumer welfare and firm profitability. The value of online variety depends on the set of products that would be available through traditional brick-and-mortar retailers in the absence of the internet. Since traditional brick-and-mortar retailers tend to cater their product assortments to local demand, we highlight the importance of accounting for

Figure 10: Increase in Retail Revenue from Local Assortments



across-market differences in demand and assortments. We build a new micro-level data set containing the sales of footwear by a large online retailer to estimate a rich model of demand allowing for consumer demand heterogeneity across markets.

The detailed nature of our data allows us to perform analysis at narrow product definitions and fine levels of geographic detail. However, it also presents us with an empirical challenge because, at these fine levels of detail, we discover an issue with small sample sizes. This is epitomized by the zeros problem, where products are observed to have zero market share. The zeros problem becomes increasingly severe at increasing levels of disaggregation, but aggregation obscures the across-market heterogeneity of interest to us. These zeros are problematic for standard demand estimation and usual remedies have been shown to generate biased estimates.

We develop new methodology to confront our small samples problem. Rather than use disaggregated local market shares directly, we use our information on location-specific sales as a type of micro-moment to augment our estimation with aggregated sales data.

Our estimation strategy exploits the structure of the model to separate the problem into two parts. At the aggregate level our estimation mimics the standard approach to pin down the demand parameters common across locations. Separately, our micro-moments are used to estimate the distribution of consumer heterogeneity across markets.

Employing our new methodology, we find products face substantial heterogeneity in demand across markets, with more niche products facing greater heterogeneity. We also show that accounting for this heterogeneity is important for rationalizing the distribution of local sales. Using our estimated model, we run a series of counterfactuals. In this analysis we find that abstracting from across-market demand heterogeneity overestimates the consumer welfare gain due to online markets by 187%. On the supply side, our estimates suggest that brick-and-mortar retail chains generate 14.5% additional revenue by localizing their assortments. Finally, we revisit the long tail phenomenon in online retail. Our results suggest that inferring consumer welfare gains from the observed aggregate long tail will tend to overstate actual welfare gains because the aggregation of sales over markets with differing demand is a key driver of the long tail.

Our approach relies on the law of large numbers in the number of markets rather than in the number of purchases. Thus, it can be useful when there are many markets and only the distribution of heterogeneity is required. In addition to measuring across-market heterogeneity, our approach is well tailored to examining the effects of discrimination by firms with knowledge of the realizations of heterogeneity. This is the context in which we apply our methodology in this paper; we could think of brick-and-mortar retailers in our application as discriminating across locations though their assortment selection. In future work, we plan to extend our methodology to include more flexible demand systems, in particular, the full random coefficients model. Additionally, we intend to apply our methodology to examine the homogenization or fragmentation of consumer tastes across regions over time.

References

Anderson, C. (2004): "The Long Tail," *Wired Magazine*, 12(10), 170–177.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81(5), 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4).

Berry, S., O. B. Linton, and A. Pakes (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *The Review of Economic Studies*, 71(3), 613–654.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25(2), 242–262.

Bronnenberg, B. J., S. K. Dhar, and J.-P. H. Dube (2009): "Brand History, Geography, and the Persistence of Brand Shares," *Journal of Political Economy*, 117(1), 87–115.

Bronnenberg, B. J., J.-P. H. Dube, and M. Gentzkow (2012): "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, 102(6), 2472–2508.

Brynjolfsson, E., Y. J. Hu, and M. S. Rahman (2009): "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science*, 55(11), 1755–1765.

Brynjolfsson, E., Y. J. Hu, and M. D. Smith (2010): "Long Tails Versus Superstars: The Effect of IT on Product Variety and Sales Concentration Patterns," *Information Systems Research*, 21(4), 736–747.

Chellappa, R., B. Konsynski, V. Sambamurthy, and S. Shivendu (2007): "An empirical study of the myths and facts of digitization in the music industry," in *Presentation 2007 Workshop Information Systems Economics (WISE), Montreal*.

Chen, W.-C. (1980): "On the Weak Form of Zipf's Law," *Journal of Applied Probability*, 17(3), 611–622.

Conlon, C. T., and J. H. Mortimer (2013): "Demand Estimation under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

Dixit, A. K., and J. E. Stiglitz (1977): "Monopolistic competition and optimum product diversity," *The American Economic Review*, pp. 297–308.

Ellison, G., and E. L. Glaeser (1997): "Geographic Concentration on U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

Gandhi, A., and J.-F. Houde (2014): "Measuring Substitution Patterns and Market Power with Differentiated Products: The Missing Instruments," Working paper, University of Wisconsin-Madison.

Gandhi, A., Z. Lu, and X. Shi (2013): "Estimating Demand for Differentiated Products with Error in Market Shares," Working paper, University of Wisconsin-Madison.

——— (2014): "Demand Estimation with Scanner Data: Revisiting the Loss-Leader Hypothesis," Working paper, University of Wisconsin-Madison.

Krugman, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of international Economics*, 9(4), 469–479.

Lancaster, K. J. (1966): "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Romer, P. (1994): "New goods, old theory, and the welfare costs of trade restrictions," *Journal of development Economics*, 43(1), 5–38.

Song, M. (2007): "Measuring consumerwelfareinthe CPU market: anapplication of the pure-characteristics demand model," *The RAND Journal of Economics*, 38(2), 429–446.

Tan, T. F., and S. Netessine (2009): "Is Tom Cruise Threatened? Using Netflix Prize Data to Examine the Long Tail of Electronic Commerce," Working paper, University of Pennsylvania, Wharton Business School.

Tucker, C., and J. Zhang (2011): "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science*, 57(5), 828–842.

Waldfogel, J. (2003): "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *RAND Journal of Economics*, 34(3), 557–568.

——— (2004): "Who Benefits Whom in Local Television Markets?," in *Brookings-Wharton Papers on Urban Economics*, ed. by J. R. Pack, and W. G. Gale, pp. 257–305. Brookings Institution Press, Washington DC.

——— (2008): "The Median Voter and the Median Consumer: Local Private Goods and Population Composition," *Journal of Urban Economics*, 63(2), 567–582.

——— (2010): "Who Benefits Whom in the Neighborhood? Demographics and Retail Product Geography," in *Agglomeration Economics*, ed. by E. L. Glaeser, pp. 181–209. University of Chicago Press, Chicago.

A An Empirical Bayesian Estimator of Shares

As mentioned in the Data section, our data exhibits a high percentage of zero observations. To account for this we implement a new procedure proposed by Gandhi, Lu, and Shi (2014). This estimator is motivated by a Laplace transformation of the empirical shares

$$s_j^{lp} = \frac{M \cdot s_j + 1}{M + J + 1}.$$

Note using that s_j^{lp} results in a consistent estimator of δ as the market size $M \to \infty$ as long as $s_j \xrightarrow{p} \pi_j$. However, instead of simply adding a sale to each product, they "propose an optimal transformation that minimizes a tight upper bound of the asymptotic mean squared error of the resulting β estimator."

The key is to back out the conditional distribution of choice probabilities, π_t, given empirical shares and market size, (s, M). Denote this condition distribution $F_{\pi|s,M}$. According to Bayes rule

$$F_{\pi|s,M}(p|s, M) = \frac{\int_{x \leq p} f_{s|\pi,M}(s|x, M) dF_{\pi|M,J}(x|M, J)}{\int_x f_{s|\pi,M}(s|x, M) dF_{\pi|M,J}(x|M, J)}.$$

Thus, $F_{\pi|s,M}$ can be estimated if the following two distributions are known or can be estimated:

1. $F_{s|\pi,M}$: the conditional distribution of s given (π, M);

2. $F_{\pi|M,J}$: the conditional distribution of π given (M, J).

$F_{s|\pi,M}$ is known from observed sales: $M \cdot s$ is drawn from a multinomial distribution with parameters (π, M),

$$M \cdot s \sim MN(\pi, M). \tag{A.1}$$

$F_{\pi|M,J}$ is not generally known and must be inferred. Gandhi, Lu, and Shi (2014) note that sales can often be described by Zipf's law, which, citing Chen (1980), can be generated if $\pi/(1 - \pi_0)$ follows a Dirichlet distribution. It is then assumed that

$$\left.\frac{\pi}{(1 - \pi_0)}\right| J, M, \pi_0 \sim Dir(\vartheta 1_J), \tag{A.2}$$

for an unknown parameter ϑ.

Equations A.1 and A.2 then imply

$$\left.\frac{s}{(1 - s_0)}\right| J, M, s_0 \sim DCM(\vartheta 1_J, M(1 - s_0)),$$

where $DCM(\cdot)$ denotes a Dirichlet compound multinomial distribution. ϑ can the be estimated by maximum likelihood, since J, M, s_0 are observed. This estimator can be interpreted as an empirical Bayesian estimator of the choice probabilities π, with a Dirichlet prior and multinomial likelihood,

$$F_{\frac{\pi}{1-s_0}|s,M} \sim Dir(\vartheta + M \cdot s).$$

For any random vector $X = (X_1, ..., X_J) \sim Dir(\vartheta)$,

$$E\left[\log(x_j)\right] = \psi(\vartheta_j) - \psi(\vartheta' 1_{d_\vartheta}),$$

Thus,

$$\begin{aligned} E\left[\log\left(\frac{\pi_j}{1-s_0}\right)\right] &= E\left[\log\left(\pi_j\right)\right] - E\left[\log\left(1-s_0\right)\right] \\ &= \psi(\vartheta + M \cdot s_j) - \psi((\vartheta + M \cdot s)' 1_{d_\vartheta}), \end{aligned}$$

which implies

$$\begin{aligned} \hat{\delta} = \log(\hat{\pi}_j) - \log(\hat{\pi}_0) &= E\left[\log\left(\pi_j\right)\right] - E\left[\log\left(\pi_0\right)\right] \\ &= \psi(\vartheta + M \cdot s_j) - \psi(M \cdot s_0). \end{aligned}$$

The nested logit model as requires an estimate of the choice probability conditional on nest,

$$\begin{aligned} \log(\hat{\pi}_j) - \log(\hat{\pi}_c) &= E\left[\log(\pi_j)\right] - E\left[\log(\pi_c)\right] \\ &= \psi(\vartheta + M \cdot s_j) - \psi\left(\sum_{j \in c} \vartheta + M \cdot s_j\right). \end{aligned}$$

B Monte Carlo Analysis

We conduct a Monte Carlo study of our proposed estimator. We start by specifying the data generating process of a nested logit demand system. We then create synthetic data sets from this process and implement the estimator.

The true model specifies consumer utility as

$$u_{i\ell j} = \underbrace{\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \xi_j}_{\delta_j} + \eta_{\ell j} + \zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}.$$

The normalized outside good gives utility $u_{i\ell 0} = \zeta_{i0} + (1-\lambda)\varepsilon_{i\ell 0}$. Here we assume both characteristics are exogenous from the unobservable ξ; however, given real data, IV methods can be used on these characteristics. We assign distributions on the data generating process according to Table 16 below.

Table 16: Data generating process

Variable	Distribution
x_1	$\mathcal{N}(0,1)$
x_2	$\mathcal{N}(0,1)$
ξ	$\mathcal{N}(0,1)$
η	$\mathcal{N}(0,\sigma=1)$
ε	T1EV
J	750
C	15
T	15
M	500000
L	100
ω_ℓ	$1/L$

Table 17 presents the results for our Monte Carlo exercises. The Berry inversion estimated using local market level share provides severely biased demand estimates. With the large number of zeros, it has a great deal of trouble identifying the nesting parameter resulting in a very noisy estimate, often exhibiting the wrong sign. Using aggregate level data provides consistent estimates of all of the parameters except for the intercept because of the nonlinear averaging of the across-market heterogeneity. Our proposed estimator performs well, while retaining the distribution of across-market demand heterogeneity.

Figure 11 compares the consumer welfare estimates under the nested logit model. Ignoring the small samples problem at the local level will overstate consumer welfare

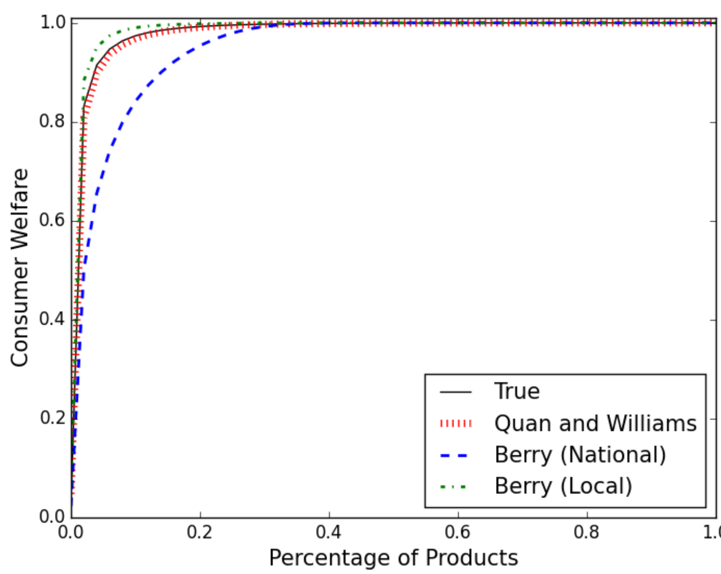
Table 17: Monte Carlo Demand Results

		Berry		Quan and
		Local	Aggregate	Williams
Parameter	True Value	(1)	(2)	(3)
β_0	-4	**-5.807**	**-3.017**	-4.125
		(1.331)	(0.475)	(0.704)
β_1	-1	**-0.719**	-1.073	-1.118
		(0.040)	(0.143)	(0.169)
β_2	0.5	**0.358**	0.535	0.557
		(0.016)	(0.072)	(0.084)
λ	0.5	**-0.490**	0.519	0.498
		(0.488)	(0.066)	(0.079)
σ	1	—	—	1.070
				(0.214)

For the Monte Carlo, we simulate 100 synthetic data sets and use the two sets of moments outlined above. The local level data averages about 71% zeros. Bold font highlights biased estimates. Standard errors in parentheses.

(dash-dot), while failing to account to the across-market demand heterogeneity will understate consumer welfare (dash). Our technique (vertical-dash), by properly accounting for both the local small sample sizes and across-market heterogeneity, is able to fit the true underlying consumer welfare (solid).

Figure 11: Consumer Welfare Estimates



Notes: Consumer welfare indexed to 1 when all products are available.