Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail

Thomas W. Quan
Department of Economics
University of Minnesota

Kevin R. Williams*
School of Management
Yale University

December 18, 2014

Abstract

This paper quantifies the effect of increased product variety in online markets on consumer welfare and firm profitability. We show the gains may be small if consumer tastes vary geographically and brick-and-mortar stores cater to the local demand. We use an original data set from a large online retailer containing millions of transactions. However, the large choice set leads to many products having zero local market shares. We propose a modification to Berry (1994) and Berry, Levinsohn, Pakes (1995), where both national and local market shares are used to recover geographically varying mean utilities. Our two step approach is easy to implement and fits our data well. Our results indicate that products face substantial heterogeneity in demand across markets, with more niche products facing greater heterogeneity. Failing to account for across-market demand heterogeneity grossly overstates the consumer welfare gain of increased online product variety, and on the supply side we find traditional retail chains can generate a substantial increase in revenue by localizing assortments.

*Authors: quanx039@umn.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, and Joel Waldfogel for useful comments. We also thank the Minnesota Supercomputing Institute (MSI) for providing computational resources.

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 effect of new products and variety on welfare either theoretically or empirically, e.g. in trade (Krugman 1979), macroeconomics (Romer 1994), and industrial organization (Dixit and Stiglitz 1977, Petrin 2002). 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 will be overstated if we do not take into account both the differences in demand across markets[1] and the fact that brick-and-mortar retailers customize their assortments to cater to local demand (Waldfogel 2010). For example, a selection of 5,000 different kinds of winter boots will be of little value to consumers living in Florida, just as a selection of 5,000 different kinds of sandals will be of little consequence to consumers in Alaska. Therefore, in order to quantify the gains from variety due to online retail, it is critical to estimate the extent to which tastes vary across regions.

We collect an extremely detailed data set consisting of point-of-sale, product review, and inventory data from a large online retailer. From July 2012 to August 2013, we observe over 13 million sales across more than 100,000 products. For each sale, we observe the date and time, shipping destination, price, and a wealth of information about the shoe. The richness of the data allows us define products and geographies at very narrow levels.

[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.

For example, we are able to differentiate between the different colors of the same model of shoe and we can attribute sales to particular metro-areas. Furthermore, as part of the project, we collected data on the assortments of Macy's and Payless ShoeSource by store. This data provides us with direct evidence that firms are responding to across-market heterogeneity, as product assortments vary significantly across stores.[2]

Using 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 highlight the extent of these differences, consider the top 1,000 products at the state level. On average, these products make up 56.5% of a state's total sales. Now consider the top 1,000 products ranked nationally. These products only make up 11.3% of a state's total sales. This suggests even among top products, demand varies significantly across locations. To formally test for differences across markets, we use simple multinomial tests that compare local market shares to national market shares. These tests 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 the estimating the gains from online variety. Our modeling approach follows the discrete choice literature with an emphasis on explicitly accounting for across-market demand heterogeneity. We allow for rich substitution patterns that are reflective of heterogeneity in tastes across locations. This will be critical for modeling the differences in a product's compensating variation across different markets. For example, removal of a popular sandal will be much more costly for markets in Florida than for

[2]Macy's, in particular, has made a concerted effort to localize product assortments. This is reflected in our data and emphasized in the following quote: "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

markets in Alaska. 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)). Failing to account for this heterogeneity will place heavy dependence on the idiosyncratic logit error, resulting in estimated welfare benefits of variety that are much too large.

Since local choice sets are often unobserved, there is the additional challenge of forecasting local choice sets for counterfactual analysis. As mentioned above, brick-and-mortar retailers tend to cater their assortments to local demand. Using our estimated demand, we infer which products local brick-and-mortar retailers would be stocking in the absence of online retail. Unfortunately, because of the number of products, the combinatorial problem of choosing the most profitable assortment of items becomes intractable. Consistent with the literature, we will assume local brick-and-mortar retailers stock the top N most popular products. This is determined by the estimated local mean utilities from the demand system.We can then calculate the compensation consumers would need to make them indifferent between a world with and without the online retailer. Notice that a model abstracting from across-market demand heterogeneity would then also assume local stores optimally stock a nationally standardized choice set, which will also lead to overestimated benefits of variety.

Our results indicate that demand for specific products varies significantly across markets, with demand for more niche products being more variable across markets. 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 18% 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 41% of the unconstrained consumer welfare is due to online variety, an overstatement of 128%. 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. Additionally, for brick-and-mortar retailers, we find a large incentive for them to cater to their local demand. By doing so they can obtain 34.7% higher revenue than under a standardized assortment.

Our results also allow us examine the effect of variety on the distribution of sales. We revisit a phenomenon called the "long tail" of online retail (Anderson 2004). The term describes a shift in the distribution of revenue toward niche, or tail, products.[3] The prevailing view is that the long tail pattern has emerged because niche products better satisfy the tastes of consumers.[4] That is, the tail is driven by consumers that switch from purchasing hit products available at their local brick-and-mortar retailers, to purchasing niche products only available online. Thus, the fact that niche products generate increasingly significant revenues has been interpreted as evidence of large welfare gains from variety.[5] For example, using a demand model that rules out systematic across-market demand heterogeneity Brynjolfsson, Hu, and Smith (2003) find large welfare gains.

Our demand model explicitly allows for systematic differences in demand across markets. This is important because across-market heterogeneity may lead to an observationally equivalent long tail. To see this consider the following example: Suppose there are 50 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 50 markets access to all 50 products. Assuming an

[3]Consider the 80/20 rule, a common rule of thumb for brick-and-mortar retailers, where 80 percent of revenue is generated by just 20 percent of products, the "hits." Put another way, niche products, the bottom 80 percent of products, account for only 20 percent revenue. However, for many online retailers niche products have been found to generate more revenue than this rule of thumb would suggest. For example, in our data, the bottom 80 percent of products accounts for 30 percent of total revenue.

[4]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.

[5]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)

equal number of consumers from each market purchase online, the online retailer will sell 50 goods that each make up 2% of sales (long tail). Therefore, inferring welfare gains from this observed long tail would be mistaken. In fact, 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. Our results suggest that, at least in our data, the long tail is primarily driven by the aggregation of sales over markets with differing demand. Additionally, in our analysis we find that an increase in variety actually reduces the share of revenue going to niche products, i.e. shortens the tail, which contradicts the current view that increased variety is the driver of the long tail.

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 over 13 million sales, the large number of products and locations, inevitably leads to many products having zero market shares. For example, even in the annualized data 59.8% of products have zero sales at the state level.[6] We could further aggregate over either geography or product space to reduce the percentage of zeros. However, the amount of aggregation required to reduce the number of zeros to negligible levels is significant and would be unsatisfactory because it would significantly smooth over the across-market heterogeneity of interest to us.

The zeros are problematic for standard demand estimation strategies because they create selection bias in the estimates (Berry, Linton, and Pakes 2004, Gandhi, Lu, and Shi 2013, Gandhi, Lu, and Shi 2014), and a contribution of our paper is to develop new methodology to address the issue. Rather than use local market shares directly to identify a product-market level fixed effect, we bring in the local market share information to form a set of micro moments that augment the aggregated (national) sales data (Petrin 2002). The differences in local market shares for products allows us to identify the variance of product-market level random effects. It is important to note that our approach estimates

[6]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.

the distribution of the heterogeneity, but not the actual realization. In this way, we can allow for estimated substitution patterns and welfare to reflect differences in the demand for products across locations. Additionally, while working with the aggregate data minimizes the zeros problem, even at the national level a few remain. We address these using a novel approach proposed by Gandhi, Lu, and Shi (2014).

Our estimation strategy exploits the structure of the 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. Separately, our micro-moments are used to estimate the distribution of consumer heterogeneity across markets, while explicitly accounting for small samples.[7] If we failed to address the small samples, we would overstate the degree of heterogeneity across markets. This will be particularly true for niche products. For example, on any given day, a niche product may sell only a single pair in the entire country. If we fail to account for the small sample issue, we might come to the conclusion that the rest of the country has absolutely no interest in the product, just because no one bought it that day. In an influential paper, Ellison and Glaeser (1997) argue that with only a small number of establishments in an industry, naive calculations will overstate the differences across locations in suitability for the industry. The same point applies when evaluating differences in demand across locations, small samples may lead to inferring a level of across-market demand heterogeneity that is spurious.[8]

The rest of the paper will be organized as follows. Section 2 discusses our data and presents preliminary evidence of across-market heterogeneity. In section 3, we present the model. Section 4 discusses our estimation procedure to be followed by our results in

[7]The key is that, given the number of observed purchases in each market, sampling from the distribution of consumer tastes implies, for each product, a certain percentage of markets will have zero sales of that product. Our estimation matches the percentage of zeros implied by simulating the model to the percentage of zeros observed in the data (our micro-moments).

[8]In a discrete choice model this will tend to overstate consumer welfare. Note that taking seriously zero observed sales when the true choice probability is greater than zero will artificially increase the choice probability of products with observed purchases. Thus, products with observed purchases will have inflated mean utilities, and hence, consumer welfare will be overstated.

Section 5. Section 6, contains our counterfactual analysis. Finally, 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 two brick-and-mortar retailers, Macy's and Payless ShoeSource. A discussion of this data can be found in Appendix A.

We begin by summarizing our data (Section 2.1), then we provide evidence of across-market consumer demand heterogeneity (Section 2.2). Finally, we document the "zeros problem" in the data and discuss aggregation as a means to address the issue (Section 2.3).

2.1 Online Shoe Sales

The main data 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. We tracked all footwear transactions between July 2012 to August 2013. 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. Each sale corresponds to a SKU (stock-keeping unit) and a numeric code for the style. The style code allows us to discern red versus blue of the same shoe model. The transaction identifier allows us to see if a customer purchased more than a single pair of shoes. For each product we record the brand, product material, and many categorical classifying variables, such as if a shoe is a wingtip and the material of the shoe. Finally, we download a picture of each shoe, and image process them to create color covariates.

We also merge in product review and inventory data. The review data contains the time series of reviews for each SKU. Each review contains reported ratings on comfort, look, and overall appeal. For the inventory data, we track daily inventory for every shoe.[9] Importantly, this data allows us to infer the total set of shoes in the consumer's choice set, even when the sale of a particular shoe is not observed.

Table 1: Online Retail Data Summary

POS Data		Total	Boys	Girls	Men	Women	Unisex
	Transactions	13.5mil	4.6%	4.4%	27.4%	63.0%	0.5%
	Price ($)	87.26	46.11	50.14	87.97	98.93	58.00
		(56.6)	(16.8)	(26.1)	(48.3)	(62.3)	(23.9)
	Products	117,493	7.1%	10.3%	25.2%	57.0%	0.5%
Transactions	*Size*	Avg	1	2	3	4	5+
		1.5	72.0%	17.9%	5.5%	2.4%	2.3%
Reviews	*Score*	Avg	1	2	3	4	5
	Comfort Rating	4.50	3.32%	3.14%	6.29%	14.32%	72.94%
	Look Rating	4.72	0.90%	1.14%	4.16%	12.52%	81.27%
	Overall Rating	4.52	2.61%	3.23%	6.84%	14.63%	72.69%
	Number of reviews		582,712				
Inventory							
	Avg daily assortment size			56,468			
	Total assortment size over year			117,493			

Table 1 provides a summary of the online retailer data sets. We observe over 13.5 million shoe transactions over the year of data collection, and over 60% of transactions are women's shoes. The price of shoes varies substantially across gender, but also within

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

gender – for example, dress shoes tend to be more expensive than walking shoes. The mean transaction size per order is 1.5 pairs of shoes; however, the distribution of transaction sizes is heavily skewed to the left. About 72% of transactions contain only a single pair, and only 2.3% of transactions contain more than five pairs. Additionally, of the transactions containing multiple purchases, only 27.3% of them (or 7.6% of all transactions) 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. In addition to the review text, we also record the consumer response to a few questions regarding the fit and look of the product. Table 1 reports the comfort, look, and overall rating across all reviews, where 1 is the lowest rating, and 5 is the highest rating. The reviews are heavily skewed towards favorable ratings, and we include this data in the demand system.

An important feature of the data is the number of products the online retailer offers. The average daily assortment size is over 56,000 products, and over the span of data collection, over 117,000 pairs of shoes were offered for sale. This constantly changing choice set provides us with additional variation that will help us identify the parameters of our model.

2.2 Across-Market Demand Heterogeneity

The premise of this paper is that there may exist significant differences in consumer demand across geographic markets. If so, we would expect local retailers to cater their inventory to their locality's consumers. This may occur through some combination of two avenues. First, while large national retailers take advantage of economies of scale through standardization, more recently many national retailers are making a push to regionally specialize their product assortments. Second, small local independent retailers are likely

to stock products based upon its local market's demand in order to compete with the larger retailers.

If our premise holds, then abstracting from heterogeneity in consumer demand across markets will overestimate the value of the increase in consumers' access to variety. The extent of this overestimation will be driven by the degree of consumer demand heterogeneity across markets, particularly for products that are highly ranked nationally. We will remain agnostic about the source of heterogeneity across markets.

Since prices, product characteristics, and choice sets are the same for all markets, differences in observed local market shares can only be rationalized by differences in local demand. In Table 2 we present the share of revenue generated by the top 1,000 products, ranked within market and ranked nationally. If demand was homogeneous across markets, we would expect the set of products composing these top 1,000 rankings to be the same and, thus, the two columns of Table 2 would be equal. Instead we see the share of revenue generated by the market level top 1,000 products is very large compared to the national top 1,000. For example, the top 1,000 products ranked at the state level make up 56.5% of revenue, but the top 1,000 products ranked nationally only accounts for 11.3% of revenue at the state level. 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

Geography	Market Top 1,000	National Top 1,000
National	0.278	0.278
State	0.565	0.113
Combined Statistical Area	0.866	0.302

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 smooths across-market heterogeneity leading in 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

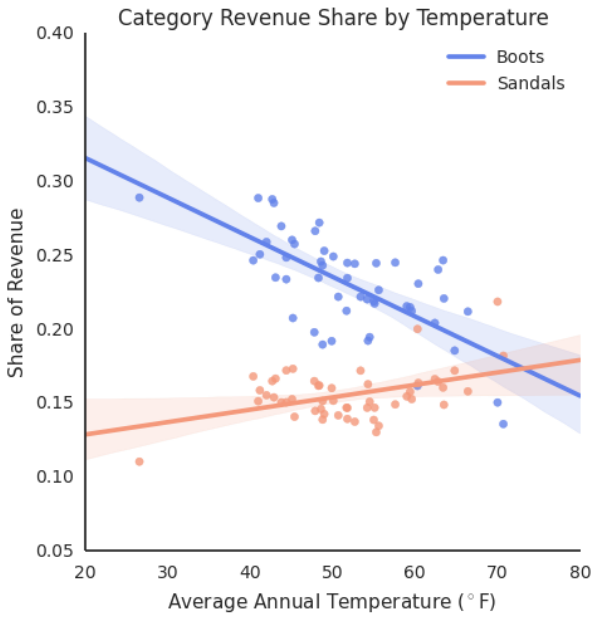
Rejection rates for multinomial tests comparing local market shares to national market shares. The null hypothesis is $H_0 : s_{\ell j} = s_j$, for all $j \in J$

Some differences across markets occur for obvious reasons. Take our earlier example of boots versus sandals, Figure 1 plots the share of state revenue captured by boots and sandals against the state's average annual temperature. As expected, boots take up a greater share of revenue in colder states and a smaller share in warmer states. Conversely, the opposite relationship holds for sandals.

Other differences across markets occur for less obvious reasons. In Figure 2, we map the consumption pattern of the most popular brand by national revenue. Annual sales are mapped into 3 digit zip codes for the eastern United States.[10] 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

[10]We isolate the eastern United States to be able to distinguish differences at fine levels of disaggregation and because the interesting portion of the map happens to be the northeastern part of the country. The full map can be viewed in Appendix C (Figure 10).

Figure 1: Share of Revenue by State: Boots vs. Sandals

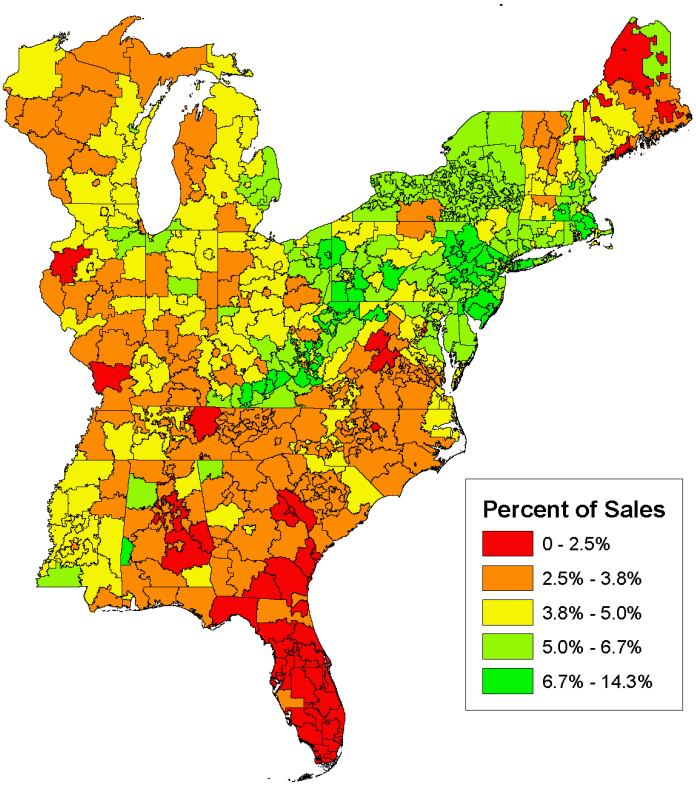


New York, New Jersey, and Massachusetts it makes up over 6% of sales. We will exploit this variation to help us identity across-market demand heterogeneity.

2.3 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 metro level (Combined Statistical Area - CSA) 95% of products have zero monthly sales. This highlights a small sample problem that is common in high frequency sale data. Observations of zero sales is problematic from both a theoretical and empirical point of

Figure 2: Sales Share of Most Popular Brand Across Zip3s



view. An in-depth discussion of these issues can be found in Berry, Linton, and Pakes (2004), Gandhi, Lu, and Shi (2013), 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 smooths 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%, but 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. 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: Product Aggregation

Product Definition	Pct. Zeros	Market Top 1,000	National Top 1,000
SKU-style	0.85	0.565	0.113
SKU	0.55	0.732	0.530
		Market Top 10	National Top 10
Brand-Category	0.51	0.311	0.299
Brand	0.30	0.382	0.368

Time horizon fixed at monthly level and geography aggregated to state level. Illustrates how product aggregation lessens burden of small sample sizes by smooths across-market heterogeneity.

The table shows a clear trade-off: At increasing levels of aggregation, the zeros problem is reduced, but this is at the expense of smoothing 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 smooth the heterogeneity in which we are interested. This motivates the need to address small sample sizes when allowing for across-market heterogeneity.

3 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, $u_{i\ell j} \geq u_{i\ell j'}, \forall j' \in J \cup \{0\}$. For a product $j \in J \cup \{0\}$, 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 in the (national) population of consumers and $\nu_{i\ell j}$ is a random utility component that is heterogeneous across consumers and locations. We decompose the random utility component into

$$\nu_{i\ell j} = \eta_{\ell j} + \varepsilon_{i\ell j},$$

where $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution and $\eta_{\ell j}$ is drawn independently from a normal distribution, $N(0, \sigma_j^2)$. These terms decompose the heterogeneity in the random utility among consumers into an "across-market" effect, $\eta_{\ell j}$, and a "within-market" effect, $\varepsilon_{i\ell j}$. The relative importance of the across-market component is determined by σ_j^2. When $\sigma_j^2 = 0$ for all $j \in J$, then the model reduces to a standard "love of variety" logit model, where there is no distinction between local and national preferences. That is, all heterogeneity is within-market heterogeneity, which is identical across locations.

For any fixed location $\ell \in L$, characterized by $\eta_\ell = \{\eta_{\ell j}\}_{j=1}^J$, we can integrate out over the within-market heterogeneity, $\varepsilon_{i\ell j}$. Since $\varepsilon_{i\ell j}$ is distributed T1EV, integrating over them forms location-specific consumer choice probabilities,

$$\pi_{\ell j} = \pi_j(\eta_\ell; \delta) = \frac{\exp\{\delta_j + \eta_{\ell j}\}}{\sum_{j'=0}^J \exp\{\delta_{j'} + \eta_{\ell j'}\}}. \tag{3.1}$$

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

$$\pi_j = \int_L \pi_j(\eta_\ell; \delta) dF\omega = \sum_{\ell=1}^{L} \omega_\ell \pi_j(\eta_\ell; \delta),$$

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

The key difficulty is that the exact location-specific fixed effects η_ℓ cannot be recovered from the sales data because of the sparsity of demand within disaggregated locations. In the next section, we outline a procedure that incorporates micro-moments – moments on disaggregated local shares – to estimate the distribution of η, essentially estimating η as a random effect. We then use traditional discrete choice tools to estimate parameters in δ. Crucially, our procedure accounts for the fact that local market share observations have small samples.

4 Estimation

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

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

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

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

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

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

where x_j is a vector of product characteristics, p_j is the price of product j, and ξ_j is the unobserved product quality for product 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[11] 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).

4.1 Micro Moments

Let $P0_{\ell j}(\sigma)$ 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,

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

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

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

[11] For example, BLP instruments

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.

$$m(\sigma) = \sum_{j=1}^{J} s_j \left(P0_j(\sigma) - \hat{P0}_j \right)^2,$$

where we weight by national market shares, s_j. We parameterize σ in the following way

$$\sigma_j = h(\log(\text{rank}_j)) = \gamma_0 + \gamma_1 \log(\text{rank}_j) + \gamma_2 \log(\text{rank}_j)^2,$$

where σ_j is allowed to depend on product j's popularity. Thus, we augment the IV moments with the micro moments $m(\sigma)$ to estimate the model parameters (γ, α, β).

Having laid the foundation of our estimation, the remaining subsections will discuss the computational mechanics. We begin by showing that our inverted choice probabilities take a convenient analytical form, which greatly simplifies the simulation of our local choice probabilities. We then show how we use this structure and the micro moments to estimate the distribution of across-market heterogeneity, σ. Finally, we discuss the identification of our parameters.

4.2 Inverting the Market Share

In this subsection, we show that the inverse of our market share 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, we believe the choice probability of the outside good, $\pi_{\ell 0}$, is well estimated in the data.[12] 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,*

[12]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. 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.

$\forall j \in J$,

$$\delta_j = \log \pi_j - \log \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell 0} \exp\{\eta_{\ell j}\}. \tag{4.1}$$

Proof. We will find it convenient to write shares as a fraction of the inside good. By Bayes rule

$$\begin{aligned}
\pi_j(\eta_\ell; \delta) &= Pr_\ell\{\text{ J }\} \cdot Pr_\ell\{\text{ j } | \text{ J }\} \\
&= (1 - \pi_{\ell 0}) \frac{\exp\{\delta_j + \eta_{\ell j}\}}{\sum_{j'=1}^{J} \exp\{\delta_{j'} + \eta_{\ell j'}\}}
\end{aligned}$$

Aggregated choice probabilities are then

$$\pi_j = \sum_{\ell=1}^{L} \omega_\ell \pi_j(\eta_\ell; \delta) = \sum_{\ell=1}^{L} \omega_\ell (1 - \pi_{\ell 0}) \frac{\exp\{\delta_j + \eta_{\ell j}\}}{\sum_{j'=1}^{J} \exp\{\delta_{j'} + \eta_{\ell j'}\}}.$$

Next, define

$$\Phi_\ell = \sum_{j'=1}^{J} \exp\{\delta_{j'} + \eta_{\ell j'}\},$$

so that $\pi_j = \sum_{\ell=1}^{L} \omega_\ell (1 - \pi_{\ell 0}) \frac{\exp\{\delta_j + \eta_{\ell j}\}}{\Phi_\ell}$. 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

$$\pi_{\ell 0} = \frac{\exp(0)}{\exp(0) + \Phi_\ell} = \frac{1}{1 + \Phi_\ell}.$$

Rewriting the equation above, in terms of Φ_ℓ, implies $\Phi_\ell = \frac{1 - \pi_{\ell 0}}{\pi_{\ell 0}}$. This expression can be substituted into the aggregate share for each inside good j, so that

$$\begin{aligned}
\pi_j &= \sum_{\ell=1}^{L} \frac{\omega_\ell (1 - \pi_{\ell 0}) \exp\{\delta_j + \eta_{\ell j}\}}{\Phi_\ell} \\
&= \exp\{\delta_j\} \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell 0} \exp\{\eta_{\ell j}\}.
\end{aligned}$$

Finally, taking logs, we then have

$$\log \pi_j = \delta_j + \log \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell 0} \exp\{\eta_{\ell j}\}$$

or

$$\delta_j = \log \pi_j - \log \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell 0} \exp\{\eta_{\ell j}\}.$$

■

Since the population shares, ω_ℓ, and the outside good shares, $\pi_{\ell 0}$, are known, this equation relates δ_j to the aggregated data, π_j. Additionally, notice that this reduces to the standard Berry (1994) inversion when $\eta_\ell = 0, \forall \ell \in L$. In the next subsection, we describe how we estimate the distribution of heterogeneity using our micro-moments. We can then integrate out this distribution to obtain the mean utilities, δ_j, from the data, π_j, and proceed with standard methods at the aggregate level.

4.3 Estimation Procedure

Local level utilities can then be written as

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

where $\bar{\eta}_{\ell j}$ is an i.i.d. draw from a standard normal distribution. For any σ, simulated local choice probabilities are then given by

$$\hat{\pi}_{\ell j} = (1 - \pi_{\ell 0}) \frac{\delta_j + \sigma_j \bar{\eta}_{\ell j}}{\sum_{j'=1}^{J} \delta_{j'} + \sigma_{j'} \bar{\eta}_{\ell j'}}.$$

The local level choice probabilities, are then used to simulate consumer purchases at each location, holding the number of observed purchases fixed. This allows us to explicitly account for small sample sizes at the location level. We then estimate $\hat{h}$ as the function that minimizes $m(\sigma)$.

After obtaining estimates of $\hat{h}$, 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 Equation 4.1.[13] We then estimate

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

using standard instrumental variables methods to control for price endogeneity. Included in x is product ratings for comfort, look, and overall, and fixed effects for color, category, 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

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

To examine the performance of our two-step estimator we perform a series Monte Carlo exercises. We find, using simulated data, that parameters are estimated precisely. A full discusion of these exercises can be found in Appendix D.

4.4 Identification

The variance of our location level random effect, $h(\cdot)$, is identified through differences in local market shares. If 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. For each product, we will use the number of locations in which zero sales are observed to form our micro moment. 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 market shares. 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.

[13] Since we take many draws over the distribution of $\eta_{\ell j}$, Proposition 2 implies that we can estimate the sum in Equation 4.1 without explicitly knowing each individual $\eta_{\ell j}$

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 η), which suggests a higher variance in the heterogeneity. This is because the higher the variance the greater the density of low η draws. Conversely, if few zeros are observed, this suggests there are few markets with low draws of η and, hence, a lower variance.

Parameters within δ_j are identified in the cross-section through variation in aggregate sales given characteristics, x_j, p_j, and across time periods through variation in the choice set J. For time varying characteristics, prices and product reviews, additional identifying power comes from intertemporal variation.

5 Results

In this section, we discuss our estimates and the fit of the model. We will define our geographic locations to be composed of 150 Combined Statistical Areas (CSAs) and our time horizons to be at the monthly level. While in our estimation it is the second step of our procedure, for exposition, we will begin by discussing the demand parameters 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.

5.1 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, category, color, and time. We also account for any remaining zeros using the correction proposed by Gandhi, Lu, and Shi (2014). A discussion of the correction procedure and results without employing the correction can be found in Appendix B.

We present four sets of estimates: (1) the logit demand model estimated at the CSA level, (2) BLP estimates at the national level, (3) our two-step estimation procedure with the distribution of across-market heterogeneity constant across products, and, our preferred specification, (4) our two-step estimation procedure allowing across-market heterogeneity to vary across products. We discuss each of these in turn.

Our first specification, the logit demand model estimated at the local level, illustrates the selection bias generated by the severity of the zeros problem. When estimating the logit model at the CSA level, each observation is a product-location specific share. Thus, the number of observations in the heterogeneous logit model is 150 times greater (number of products times 150 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 attenuated. Of particular concern for us are the price coefficients, which are attenuated by an order of magnitude, compared to our other specifications. In the bottom panels of each table, we can see that this specification implies price elasticities that are much too inelastic, ten times smaller than our other specifications. This, in turn, will imply consumer surplus estimates that are much too high.

We use specifications (2) and (3) to directly compare results estimated using standard approaches and results estimated using our procedure. There is a subtle difference between the two specifications. In the BLP estimation, the random coefficient corresponds to an individual drawn from the national population, while in our estimation the random coefficient corresponds with a location. Unsurprisingly, the results for these specifications

Table 6: Demand Estimates - Men's

	Local Logit (1)	National BLP (2)	Homoskedastic 2-Step (3)	Heteroskedastic 2-Step (4)
Price	-0.014	-0.103	-0.107	-0.117
	(0.000)	(0.000)	(0.007)	(0.008)
Comfort	0.043	0.181	0.192	0.214
	(0.004)	(0.000)	(0.043)	(0.047)
Look	-0.108	-0.704	-0.717	-0.778
	(0.004)	(0.000)	(0.059)	(0.064)
Overall	0.180	0.800	0.813	0.886
	(0.005)	(0.000)	(0.056)	(0.061)
No Reviews	0.339	2.906	3.003	3.321
	(0.013)	(0.355)	(0.284)	(0.311)
Constant	-13.283	-10.552	-9.191	-8.956
	(0.030)	(0.004)	(0.627)	(0.690)
σ	—	1.089	1.011	$*$
		(0.001)		
Fixed Effects				
Brand	✓	✓	✓	✓
Category	✓	✓	✓	✓
Color	✓	✓	✓	✓
Month	✓	✓	✓	✓
N	1,273,124	164,241	164,241	164,241
Zeros	23,363,026	14,974	14,974	14,974
	(94%)	(9%)	(9%)	(9%)
Price Elast.				
Product	-1.271	-11.723	-12.100	-13.226
	(0.726)	(8.683)	(8.962)	(9.800)
Industry	-0.010	-0.110	-0.088	-0.094

Notes: Estimated at the monthly level. "Local Logit" (1) estimates the logit model at the CSA level, hence the ξ's are market level fixed effects. "National BLP" (2) estimates the model with the BLP contraction at the national level. Finally, we report our two-step procedure allowing for across-market heterogeneity to be constant across products (3) and to vary across products (4).

All reported coefficients are significant at the 1% level.

$*$ estimates for across-market heterogeneity in specification (4) will be discussed in the following subsection.

Table 7: Demand Estimates - Women's

	Local Logit (1)	National BLP (2)	Homoskedastic 2-Step (3)	Heteroskedastic 2-Step (4)
Price	-0.001	-0.010	-0.011	-0.012
	(0.000)	(0.005)	(0.008)	(0.001)
Comfort	0.048	0.015	0.023	0.028
	(0.003)	(0.003)	(0.008)	(0.008)
Look	-0.069	-0.221	-0.225	-0.242
	(0.002)	(0.020)	(0.007)	(0.008)
Overall	0.111	0.269	0.271	0.299
	(0.003)	(0.022)	(0.010)	(0.010)
No Reviews	0.036	-0.194	-0.151	-0.128
	(0.007)	(0.246)	(0.039)	(0.042)
Constant	-14.158	-17.759	-16.956	-17.422
	(0.020)	(0.362)	(0.064)	(0.070)
σ	—	1.106	1.191	*
		(0.001)		
Fixed Effects				
Brand	✓	✓	✓	✓
Category	✓	✓	✓	✓
Color	✓	✓	✓	✓
Month	✓	✓	✓	✓
N	2,448,538	328,598	328,598	328,598
Zeros	46,841,162	34,831	34,831	34,831
	(95%)	(10.5%)	(10.5%)	(10.5%)
Price Elast.				
Product	-0.113	-1.241	-1.306	-1.405
	(0.070)	(1.069)	(1.125)	(1.210)
Industry	-0.001	-0.010	-0.010	-0.011

Notes: Estimated at the monthly level. "Local Logit" (1) estimates the logit model at the CSA level, hence the ξ's are market level fixed effects. "National BLP" (2) estimates the model with the BLP contraction at the national level. Finally, we report our two-step procedure allowing for across-market heterogeneity to be constant across products (3) and to vary across products (4).

All reported coefficients are significant at the 1% level.

$*$ estimates for across-market heterogeneity in specification (4) will be discussed in the following subsection.

are very similar. However, the advantage to our approach is that it estimates the distribution of heterogeneity across locations, rather than across individuals. The importance of this distinction will be highlighted in the follwoing section when we do counterfactual analysis at the location level.

We now turn to our preferred estimates, specification (4) allowing for across-market heterogeneity to vary across products. The price coefficients have the expected signs, -0.117 and -0.012 for men's and women's shoes, respectively. These results suggest that men are far more price sensitive (-13.226) than women (-1.405) when it comes to their footwear purchases. Turning to the coefficients on our review variables, we can see that the comfort and overall ratings have the expected sign, with higher ratings having positive effects on demand. Look, however, appears to have an opposite sign than expected. Upon closer examination of our product ratings, it appears that the rating for look is often higher than the ratings for comfort and overall appeal. Perhaps the qualities that make a shoe aesthetically pleasing reduces its appeal through other channels. Our indicator for no reviews takes on opposite signs for men's and women's shoes. This variable largely captures the demand for new products. The composition of sales provides some insight into the differing effects by gender. Sales of men's shoes are concentrated in sneakers, while sales of women's shoes are more concentrated toward boots, heels, and sandals. It may be that sneakers are a more standardized items lessening the importance of review information.

Comparing our preferred specification to specification (3), we again see that the parameters constant across markets are quite similar, but they are slightly greater in magnitude for our preferred specification. In the next section, we will show that the additional flexibility of allowing across-market heterogeneity to vary by product will be important to rationalizing the distribution of local sales. This suggests that failing to allow for this flexibility in specification (3) may introduce measurement error into the inverted δ's resulting in a small attenuation bias.

5.2 Across-Market Heterogeneity

Our results in the previous subsection depended on our estimate of $h(\cdot)$, the computation of which we expand upon here. We estimate the distribution of across-market heterogeneity

$$\sigma_j = h(\log(\text{rank}_j)) = \gamma_0 + \gamma_1 \log(\text{rank}_j) + \gamma_2 \log(\text{rank}_j)^2,$$

by minimizing the sum of squared errors on the products' percentage of locations with zero sales, weighted by observed national sales. Our estimates for the full specification and for the specification with σ_j constant across products, i.e. $h(\cdot) = \gamma_0$, are presented in Table 8.

In the full specification, corresponding to our demand estimates in specification (4), we can see that σ_j is increasing as popularity decreases. To get a sense of the magnitude of this heterogeneity, we also report the range and standard deviations of the resulting δ_j estimates. The heterogeneity, particularly for lower ranked products is quite large, approaching the standard deviation observed in the estimated mean utilities. This suggests products that are unpopular, on average, may be very popular in particular markets. Since we weight our objective function by observed sales, the σ_j we estimate in specification (3) is closer to the estimated heterogeneity of the most popular products in the full specification.

Figure 3 gives us further insight into our heterogeneity results and illustrates how well our first stage estimation fits. 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. The bottom panels zooms into the top 20,000 observations. For comparison, we include simulation results for the case of homogeneous demand across markets, i.e. when $\sigma_j = 0$. 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 increases to account for the increasing percentage of zero market shares. If demand were homogeneous across

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

	Men		Women	
	(3)	(4)	(3)	(4)
γ_0	1.011	0.647	1.191	0.721
γ_1		0.092		0.091
γ_2		0.001		0.001
SSE	1,434	1,354	2,563	2,495
N	164,241		328,598	
Product Rank		σ_j		σ_j
100		1.094		1.164
1,000		1.335		1.404
15,000		1.633		1.700
δ_j				
Range	14.038		15.123	
St. Dev.	1.858		1.941	

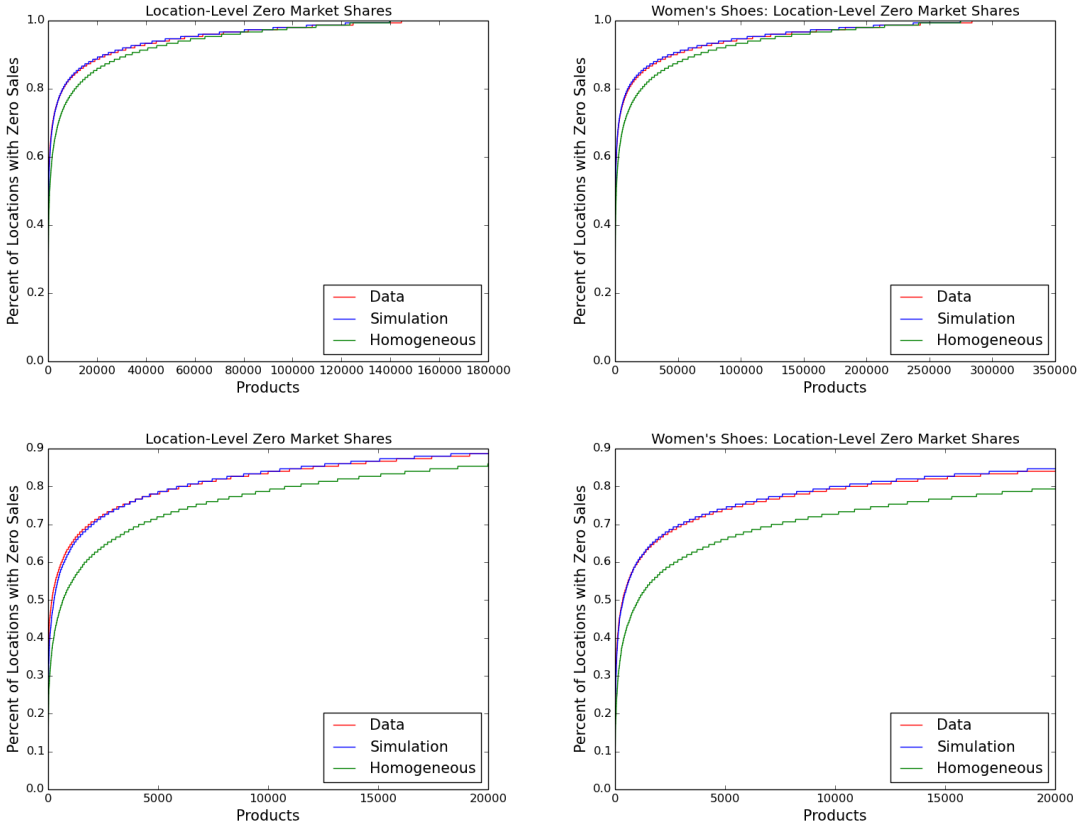
Two step results for the distribution of across-market heterogeneity. Specification (3) restricts the variance of the across-market heterogeneity to be constant across products, while specification (4) allows the variance vary by popularity. The bottom panel presents summary information on δ for comparisions of magnitudes.

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

6 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 being by presenting our primary results, allowing for tastes to differ across markets and for local brick-and-mortar retailers to cater their assortments to local demand (Section 6.1). We will then present results shutting down across-market heterogeneity and show how these results overestimate the gains to online

Figure 3: 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 (red), in our estimation with across-market heterogeneity (blue), and with homogeneous demand across markets (green).

variety (Section 6.2). Finally, we revisit the phenomenon of the long tail and show that aggregation of sales over markets with different tastes is the primary driver of the long tail of online retail (Section 6.3).

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-mortars stock the top N most popular products. This is determined by the estimated local mean utilities from the demand system. Notice that

a model assuming common national tastes would then also assume local stores optimally stock a nationally standardized choice set. While the literature often establishes the same threshold for all markets, 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 benchmark for our selection of local level assortment sizes. We will report results for both a range of thresholds and for our benchmark.

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 3.1. Using these probabilities we simulate location level purchases which then allows us to compute counterfactual consumer surpluses and retail revenues.

6.1 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 surplus

$$CS_\ell = \frac{M\omega_\ell}{\alpha} \log\left(1 + \sum_{j=1}^{J} \exp\{\delta_j + \eta_{\ell j}\}\right),$$

and retail revenue,

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

where M is the national population size. Table 9 and Table 10 present the consumer surplus and retail revenue, respectively, under restricted choice sets relative to the unconstrained choice set for each of our specifications.

The deficiencies of the alternative specifications are highlighted when compared to our preferred specification, the heteroskedastic two step estimator. Employing a local level logit tends to overstate heterogeneity across markets by assuming products without an observed sale are completely unwanted at that particular location. Thus, there will be a tendency to overestimate the consumer surplus generated by products with observed sales. While this is difficult to see in the ratios, since the price coefficient cancels out, a comparison of the absolute consumer surplus shows that the estimated consumer surplus is much too high. The homoskedastic two step estimator understates across-market heterogeneity and, hence, consumer welfare. This arises because the homoskedastic specification cannot rationalize higher across-market heterogeneity for lower ranked products. Note that we omit the national BLP specification. While this specification may be consistent with across-market demand heterogeneity, there is no way to determine the underlying geographic distribution of heterogeneity.

Table 9: Localized Choice Set: Consumer Welfare - Share of Unconstrained

Assortment Size	Local Logit	Homoskedastic 2-Step	Heteroskedastic 2-Step
Benchmark	0.85	0.73	0.82
Threshold			
3,000	0.76	0.64	0.72
6,000	0.88	0.78	0.85
12,000	0.97	0.91	0.94
24,000	1	0.98	0.99
Unconstrained ($\sim 48,800$)	1	1	1
Absolute ($ Millions)	4,289.6	461.0	422.0

Examining the results of our preferred specification, we can see the gain to additional variety is fairly small when local stores cater to local demand. If local stores stock just 3,000

well targeted products, consumers would capture 72% of the unconstrained consumer surplus, the total consumer surplus they would obtain with access to all of the products. At the benchmark assortment size,[14] consumers would capture 82% of the unconstrained consumer welfare. Conversely, only 18% of the unconstrained consumer welfare is due to access to online variety. Similar conclusions can be drawn for retailer revenue. A national brick-and-mortar chain can generate 69% of the total revenue it would generate by stocking the universe of products, by stocking just 3,000 well selected products.

Table 10: Localized Choice Set: Retail Revenue - Share of Unconstrained

Assortment Size	Local Logit	Homoskedastic 2-Step	Heteroskedastic 2-Step
Benchmark	0.71	0.55	0.66
Threshold			
3,000	0.73	0.61	0.69
6,000	0.88	0.76	0.83
12,000	0.97	0.90	0.94
24,000	1	0.98	0.99
Unconstrained ($\sim 48,800$)	1	1	1
Absolute ($ Millions)	687.0	687.0	687.0

6.2 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 same subset of top products determined by ranking products according to their national mean utilities, δ_j. We will use our BLP estimates with the assumption that consumer types are evenly dispersed across locations, as the basis of our comparisons with the previous subsection.

[14] The number of products held by Macy's and Payless in each CSA.

Table 11 and Table 12 present the consumer surplus and retail revenue, respectively, under nationally standardized restricted choice sets relative to the unconstrained choice set.

Table 11: National Choice Set: Consumer Welfare - Share of Unconstrained

Assortment Size	Localized Choice Set	National BLP	Heterosked. 2-Step
Benchmark	0.82	0.59	0.60
Threshold			
3,000	0.72	0.49	0.49
6,000	0.85	0.65	0.65
12,000	0.94	0.82	0.83
24,000	0.99	0.96	0.96
Unconstrained ($\sim 48,800$)	1	1	1
Absolute ($ Millions)	422.0	513.6	422.0

When choice sets are nationally standardized consumers capture much less of the total unconstrained consumer surplus. If each market is constrained to 3,000 of the most popular national products, consumers would capture 49% of the unconstrained consumer surplus and retailers would capture 46% of the unconstrained revenue. Notice that our results are nearly identical, whether the model is estimated using BLP or our two step method. The exception is that the absolute consumer surplus is slightly high under the BLP estimates. 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. However, because the estimated price coefficients are slight smaller in magnitude for BLP, we estimate a slightly higher consumer welfare.

Ultimately, failing to account for localization will overestimate gains to consumer surplus when moving from a constrained choice set to the entire choice set. At the benchmark assortment size, under a standardized national assortment 41% [1-0.59] of the

Table 12: National Choice Set: Retail Revenue - Share of Unconstrained

Assortment Size	Localized Choice Set	National BLP	Heterosked. 2-Step
Benchmark	0.66	0.49	0.49
Threshold			
3,000	0.69	0.46	0.46
6,000	0.83	0.62	0.62
12,000	0.94	0.80	0.80
24,000	0.99	0.95	0.95
Unconstrained ($\sim 48,800$)	1	1	1
Absolute ($ Millions)	687.0	687.0	687.0

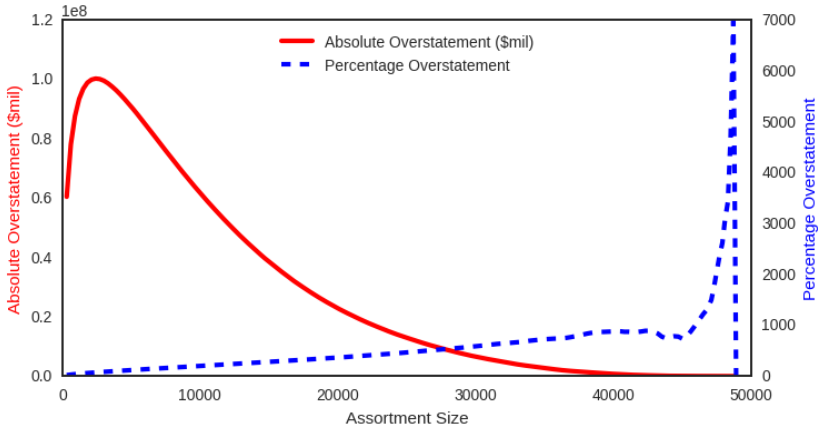
unconstrained consumer welfare is due to access to online variety, but with localized assortments the portion is just 18% [1-.82]. This suggests failing to account for heterogeneity across markets will overestimate consumer welfare due to online variety by 128%.

Similar conclusions can be drawn for retailer revenue. A researcher assuming demand is homogeneous across markets will severely underestimate counterfactual brick-and-mortar revenue. A national brick-and-mortar chain would generate just 46% of unconstrained revenues if it did not localize its assortment. This suggests that there is a significant incentive for local stores to cater to local tastes. By doing so they would obtain 34.7% greater revenue than under a nationally standarized assortment.

Figure 4 plots the estimated consumer welfare overstatement when assuming no localization, measured in millions of dollars (blue) and as a percentage (red). The absolute overstatement peaks at $100.3 million with about 2,400 products. The percentage overstatement is largest toward the tail of the distribution. Initially, this rise in percentage overstatement is driven by the increasing gap in consumer welfare between the across-market heterogeneity and across-market homogeneity counterfactuals. However, on the

right half of the distribution, consumer welfare gains are tiny after accounting for across-market heterogeneity making the percentage increase large despite the absolute levels of the gap being relatively small.

Figure 4: Overestimation of Consumer Surplus

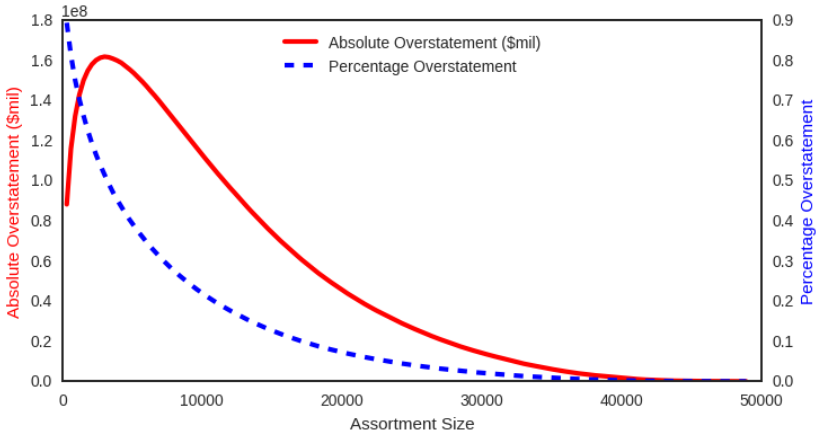


Similarly, Figure 5 graphs the increase in retail revenue due to localization of assortments, measured in millions of dollars (blue) and as a percentage (red). The absolute gain in revenue from localization peaks at $161.9 million at 3,000 products. The percentage gain is monotonically decreasing with assortment size. The graph shows that when assortment sizes are extremely limited, brick-and-mortar retailers can significantly boost revenue by maintaining localized product assortments.

6.3 Long Tail Analysis

In our analysis of the long tail we seek to answer three questions. First, in the absence of an online retailer, what would the revenue distributions look like at the local level? Second, summing these sales across locations, what would the revenue distribution look like at the national level? Finally, how does the national revenue distribution change when an online retailer enters and gives consumers access to the universe of products? The first two

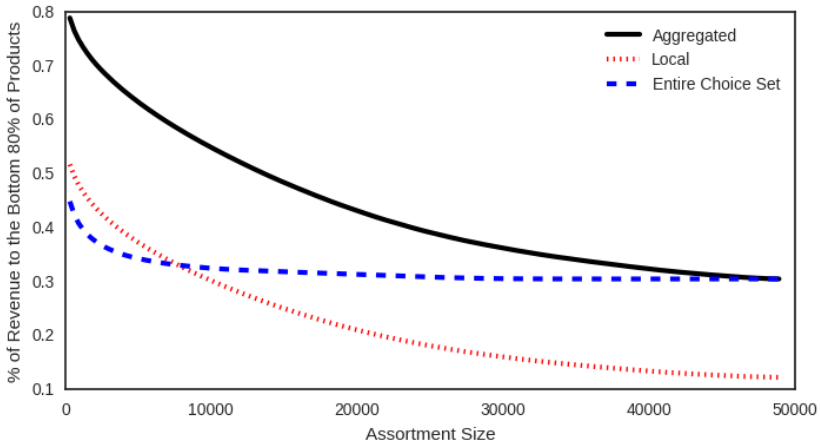
Figure 5: Additional Revenue from Regional Specialization



questions concern a hypothetical world without online retail. To answer these questions, we use our primary set of counterfactuals, in which local brick-and-mortar retailers are limited by the number of products that they can stock, but are able to select products based upon local consumer demand.

Figure 6 illustrates the decomposition of the long tail over a range of values for the assortment size threshold. The red dotted line denotes the average share of revenue going to tail products under a restricted choice set at the location level. After aggregating these sales across locations and re-ranking products based on national sales, the black solid line represents the share of revenue accruing to tail products at the national level. We can see that the share of revenue attributable to tail products is greater at the national level. That is, aggregation of sales across markets (without any product switching) lengthens the revenue tail. This is due to the fact that popularity of products varies wildly across geographic markets. As the assortment size restriction is relaxed, revenues become more concentrated toward the head, so the revenue tail becomes shorter at both the aggregated and local levels, but the lengthening effect of aggregation persists throughout the entire range.

Figure 6: Long Tail Decomposition



We then allow access to the universe of products, under the counterfactual restriction, by all consumers. This revenue distribution at the national level is represented with the blue dashed line. It shows that, in fact, access to variety serves to shorten, not lengthen, the tail as the prevailing long tail theory would imply. Therefore, we find that the lengthening of the tail in online shoe retail is primarily driven by across-market demand heterogeneity. Consumers in different markets demand different products causing a flattening effect on the distribution of revenue when measured at the national level. When viewed from the local level, however, the revenue distribution continues to exhibit a short tail. These findings suggest that consumer welfare gains inferred from an observed long tail may be overstated.

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 available through traditional brick-and-mortar retailers. Since traditional brick-

and-mortar retailers tend to cater their product assortments to local demand, we highlight the importance of accounting for across-market demand heterogeneity. 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 smooths over the across-market heterogeneity in which we are interested. 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 128%. On the supply side, our estimates suggest that brick-and-mortar retail chains generate 34.7% 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 long tail will tend to overstate actual welfare gains. Additionally, we find that an increase in variety actually shortens the tail, which contradicts the prevailing view that increased variety is the 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, for example nested logit and full random coefficients. 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. D. Smith (2003): "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49(11), 1580–1596.

——— (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.

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.

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 Localization in Footwear Retail

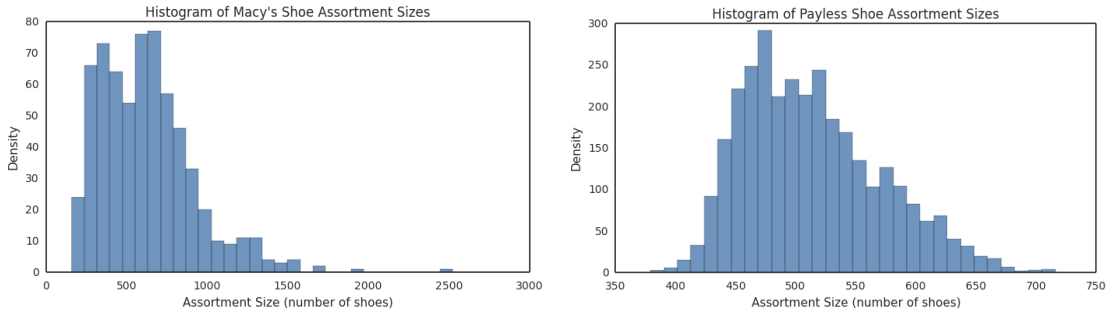
In addition to the retail data, we collect a snapshot of shoe availability for Macy's and Payless ShoeSource during August and September of 2014. We first collected all the shoe SKUs each retailer sold, and then for each SKU, we used the firm's "check in stores" web feature to see if the product was 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. 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 13: 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 13 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 which 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 7 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.

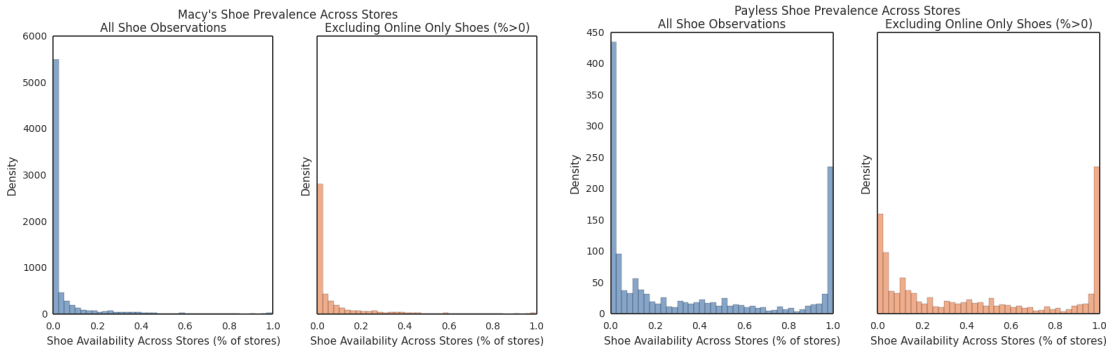
We want to measure how assortments vary by store. Figure 8 graphs the percentage of locations carrying a shoe style for Macy's and Payless. That is, we present a histogram of shoe presence across stores of the chain. If all shoes were available at all stores, the density would collapse at 1 (100%). The level panels within a chain plots the density for all shoes, whereas the right panel 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 and almost all



(a) Macy's (b) Payless Shoes

Figure 7: Shoe Assortment Size Distributions Across Retail Chains

stores. In recent years, Macy's has made a concerted effort to better localize their product assortments through a program called "My Macy's."[15] 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

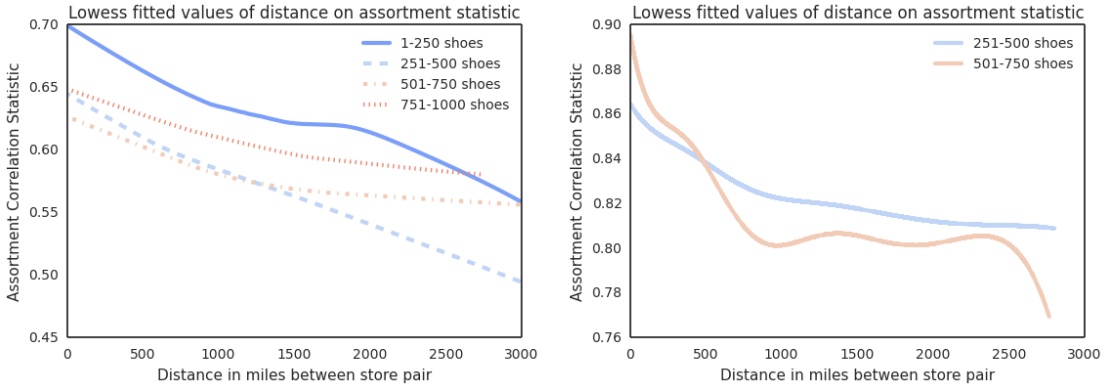
Figure 8: Footwear Prevalence Across Stores

[15] "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`

Finally, 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

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

This measure is bounded between zero and one. We use the minimum cardinality, rather than the cardinality of the union, 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 9 plots this exercise for Macy's and Payless as a function of distance between stores A and B.



(a) Macy's

(b) Payless

Figure 9: 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.

B 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),

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

$F_{\pi|M,J}$ 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

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

for an unknown parameter ϑ.

Equations B.1 and B.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

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

B.1 Comparison of Results with and without Correction

We compare our two-step procedure with alternative estimation procedures, such as accounting for endogeneity in prices as well as addressing market shares are measured with error. In Table 14 and Table 15, Local corresponds to our two-step procedure, National indicates a procedure abstracting from across market heterogeneity, AS indicates adjusted shares, and ES indicates using empirical shares, where shares equal to zero are dropped from the analysis. Finally, we indicate accounting for endogeneity in prices by IV and OLS.

Table 14: Alternative demand specification results of men's shoes

Men's Shoes	Local AS IV	Local AS OLS	National AS IV	National ES IV	National AS OLS	National ES OLS
Price	-0.117	-0.004	-0.107	-0.087	-0.004	-0.003
	(0.008)	(0.000)	(0.007)	(0.007)	(0.000)	(0.000)
Comfort	0.214	0.011	0.190	0.144	0.005	0.027
	(0.047)	(0.011)	(0.043)	(0.034)	(0.010)	(0.008)
Look	-0.778	-0.214	-0.719	-0.525	-0.203	-0.134
	(0.064)	(0.011)	(0.059)	(0.047)	(0.010)	(0.008)
Overall	0.886	0.342	0.816	0.668	0.319	0.259
	(0.061)	(0.012)	(0.056)	(0.048)	(0.011)	(0.009)
No Reviews	3.321	-0.228	2.996	2.185	-0.246	-0.049
	(0.311)	(0.035)	(0.284)	(0.214)	(0.032)	(0.025)
Constant	-8.956	-17.093	-8.685	-10.452	-16.126	-16.200
	(0.690)	(0.057)	(0.626)	(0.536)	(0.052)	(0.039)
Fixed Effects						
Brand	✓	✓	✓	✓	✓	✓
Category	✓	✓	✓	✓	✓	✓
Color	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓
N	164,241	164,241	164,241	149,267	164,241	149,267
Price Elast.						
Product	-13.226	-0.479	-12.102	-9.549	-0.455	-0.360
	(9.800)	(0.354)	(8.960)	(6.641)	(0.337)	(0.250)
Industry	-0.094	-0.004	-0.088	-0.073	-0.004	-0.003

Data aggregated to the monthly level. AS: adjusted shares; ES: empirical shares.

Table 15: Alternative demand specification results of women's shoes

Women's Shoes	Local AS IV	Local AS OLS	National AS IV	National ES IV	National AS OLS	National ES OLS
Price	-0.012	-0.004	-0.011	-0.003	-0.004	-0.003
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Comfort	0.028	0.017	0.023	0.043	0.012	0.043
	(0.008)	(0.008)	(0.008)	(0.005)	(0.007)	(0.005)
Look	-0.242	-0.203	-0.226	-0.139	-0.189	-0.139
	(0.008)	(0.007)	(0.007)	(0.005)	(0.007)	(0.005)
Overall	0.299	0.260	0.272	0.220	0.235	0.220
	(0.010)	(0.009)	(0.010)	(0.007)	(0.009)	(0.007)
No Reviews	-0.128	-0.519	-0.147	-0.222	-0.515	-0.222
	(0.042)	(0.022)	(0.039)	(0.015)	(0.020)	(0.015)
Constant	-17.422	-16.567	-16.270	-16.317	-15.519	-16.317
	(0.070)	(0.048)	(0.064)	(0.032)	(0.044)	(0.032)
Fixed Effects						
Brand	✓	✓	✓	✓	✓	✓
Category	✓	✓	✓	✓	✓	✓
Color	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓
N	328,598	328,598	328,598	293,767	328,598	293,767
Price Elast.						
Product	-1.405	-0.479	-1.326	-0.388	-0.452	-0.361
	(1.210)	(0.412)	(1.140)	(0.305)	(0.390)	(0.284)
Industry	-0.011	-0.004	-0.011	-0.003	-0.004	-0.003

Data aggregated to the monthly level. AS: adjusted shares; ES: empirical shares.

C Additional Tables and Figures

Figure 10: Sales Share of Most Popular Brand Across Zip3s

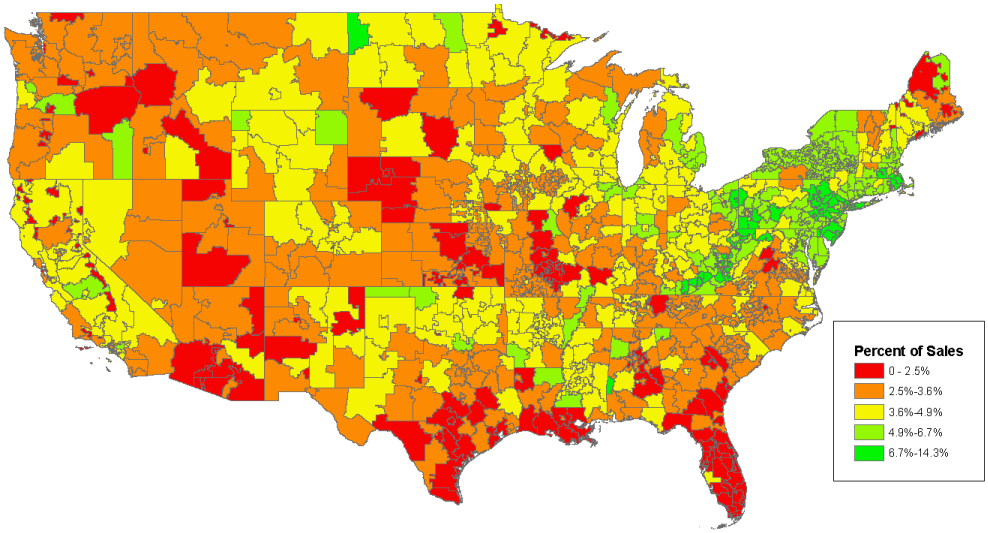
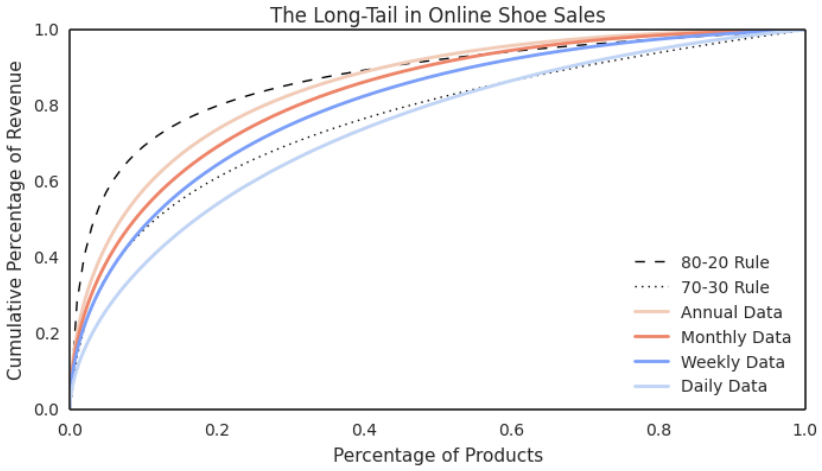


Figure 11: Long tail in the data at different levels of aggregation



D Monte Carlo Analysis

In this section, we conduct a Monte Carlo study of the two-step procedure, where local shares are used to estimate parameters governing across-market heterogeneity and aggregate shares are used to estimate parameters constant across markets. We start by assigning parameters and drawing consumer purchases from disaggregated local shares. The true model specifies

$$u_{ij\ell} = \underbrace{.5 - .5x_{1j} + 1x_{2j} + \xi_j}_{\delta_j} + \eta_{j\ell} + \varepsilon_{ij\ell}.$$

The outside good gives utility $u_{i0\ell} = \varepsilon_{i0\ell}$. We assign the following distributions on the data generating process:

Table 16: Data generating process distributions

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	500
T	1
M	500000
L	500
ω_ℓ	$1/L$

This is a special case of the empirical application where $h(\cdot)$ is homoskedastic, i.e. $\sigma_j = 1 \forall j \in J$. With the synthetic data, we have a a matrix of local shares across products and locations, $\mathbf{s}_{JL}$. Demand at each locality is obtained from simulating $\lfloor \omega_\ell M \rfloor$ consumer purchases (of $J \cup \{0\}$) according to probabilities $(\mathbf{s}_{J\ell}, 1 - \sum_J s_{j\ell})$. This DGP gives roughly 55% zeros at the local level.

The estimation routine has two steps:

1. Estimate the parameters governing η. For the Monte Carlo, this corresponds to estimating the single paramter σ. The micro-moments we use are the number locations with zero shares. We use the Nelder-Mead method to estimate σ.

2. Estimate the mean utility parameters given the estimate of $h(\cdot)$, and hence estimate of δ. Here we just estimate

$$\delta = X\beta + \xi$$

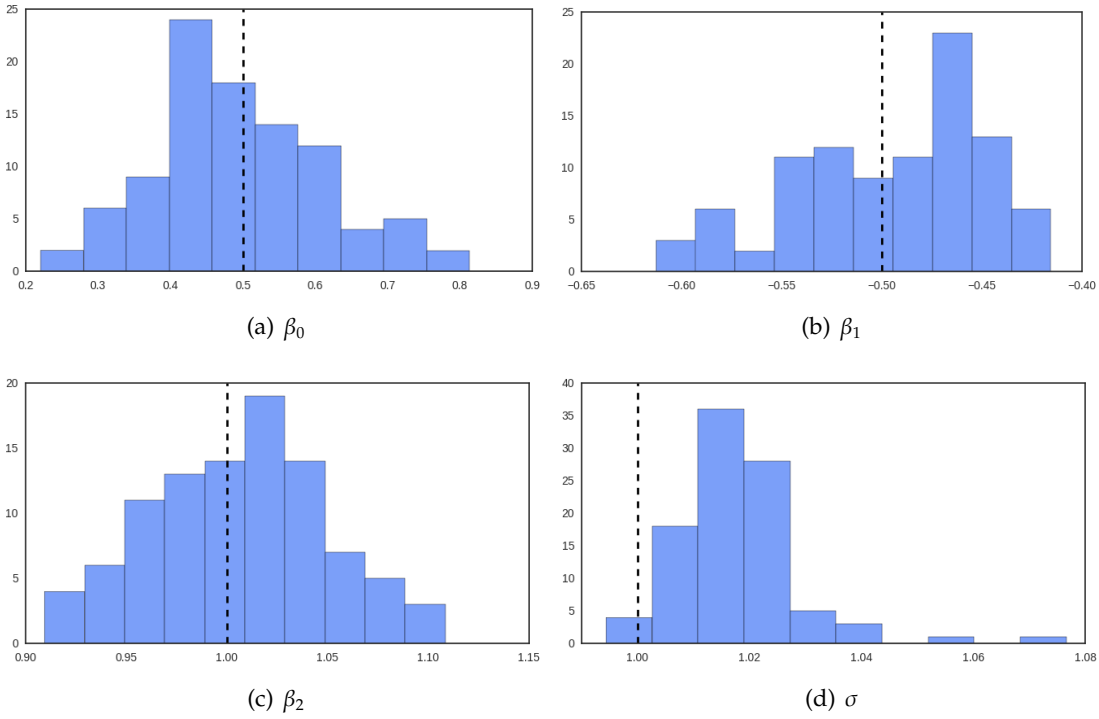
using linear regression (or using IV methods if covariates are endogenous).

Table 17 shows descriptive statistics for the Monte Carlo study. Figure 12 shows histograms of the parameter estimates. As both the table and figure show, the parameters are estimated precisely, with small mean-squared errors.

Table 17: Monte Carlo Results

Parameter	True Value	Bias	MSE
β_0	0.5	0.0067	0.0138
β_1	-0.5	-0.0039	0.0021
β_2	1	-0.0061	0.0019
σ	1	-0.0176	0.0004

For the Monte Carlo, we simulate 96 synthetic data sets and implement the two-step procedure outlined above. On a 24-core machine at 3.5GHz, the Monte Carlo takes less than 15 minutes to run.



(a) β_0 (b) β_1

(c) β_2 (d) σ

Figure 12: Histograms of parameter estimates for Monte Carlo study