11-1-2016

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

Thomas W. Quan

Kevin R. Williams

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

By

Thomas W. Quan and Kevin R. Williams

November 2016
Revised June 2018

COWLES FOUNDATION DISCUSSION PAPER NO. 2054R3

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

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

Thomas W. Quan* Kevin R. Williams†

June 2018‡

Abstract

Online retail gives consumers access to an astonishing variety of products. However, the additional value created by this variety depends on the extent to which local retailers already satisfy local demand. To quantify the gains and account for local demand, we use detailed data from an online retailer and propose methodology to address a common issue in such data—sparsity of local sales due to sampling and a significant number of local zeros. Our estimates indicate products face substantial demand heterogeneity across markets; as a result, we find gains from online variety that are 45% lower than previous studies.

JEL Classification: C13, L67, L81

*University of Georgia; quan@uga.edu

†Yale University; kevin.williams@yale.edu

‡We thank Pat Bajari, Tat Chan, Ying Fan, Ben Shiller, and Catherine Tucker for discussing this article at conferences. We thank Steve Berry, Judy Chevalier, Thomas Holmes, Kyoo il Kim, Amil Petrin, Fiona Scott Morton, Joel Waldfogel, and especially Chris Conlon, Amit Gandhi, Aniko Öry, and Mykhaylo Shkolnikov for useful comments. We also thank the participants at numerous conferences and seminars. In 2015: Darmouth-Tuck Winter IO, Northeastern University, Washington University at St. Louis, Michigan State University, Department of Justice, Bureau of Labor Statistics , Federal Communications Commission, University of Georgia, University of California at Los Angeles, NBER Winter Digitization Meeting, Singapore Management University, National University of Singapore, Stanford University (Marketing and Economics), IIOC, University of Maryland, University of Toronto, NYU Stern IO Day, QME. In 2016: Harvard University, MIT, Boston College, NBER IO Summer Institute. In 2017: the Federal Reserve Bank of Minneapolis. Finally, we thank the Minnesota Supercomputing Institute (MSI) for providing computational resources, Mauricio Cáceres Bravo for his excellent research assistance, and the online retailer for allowing us to collect the data used in this study.

1 Introduction

There is widespread recognition that as economies advance, consumers benefit from 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, Broda and Weinstein 2006, Arkolakis et al. 2008), macroeconomics (Romer 1994), and industrial organization (Lancaster 1966, Dixit and Stiglitz 1977, Brynjolfsson, Hu, and Smith 2003). 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 a dramatic increase in product variety contribute to welfare?

The central idea of this article is that the gains from online variety depend critically on the extent to which demand varies across geographies and on how traditional brick-and-mortar retailers respond to those local tastes (Waldfogel 2008, Waldfogel 2010, Thomas 2011).[1] For example, online access to an additional 5,000 different kinds of winter boots will be of little value to consumers living in Florida, just as access to an additional 5,000 different kinds of sandals will be of little consequence to consumers in Alaska. If Alaskan retailers already offer a large selection of boots that captures the majority of local demand, only consumers with niche tastes – possibly those who have a trip planned to Florida – will benefit from the additional 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 both within and across locations.

Influential work by Brynjolfsson, Hu, and Smith (2003) found significant gains to consumer welfare ($731 million - $1.03 billion in 2000) due to the increase in access to book varieties provided by Amazon.com. They estimate the gain to consumers from increased variety to be seven to ten times larger than the gain derived from the competitive price effect. These gains have since been dubbed the "long tail" benefit of online retail by Anderson (2004). This term captures the idea that although the value of each additional variety individually contributes only a tiny fraction to consumer welfare, the summation of an enormous number of small gains becomes sizable. However, the existing literature has been limited by the aggregate nature of their data and have been forced to abstract from local heterogeneity.

[1] A large body of literature that has highlighted across-market differences in demand, including Waldfogel (2003), Waldfogel (2004), Bronnenberg, Dhar, and Dube (2009), Choi and Bell (2011), and Bronnenberg, Dube, and Gentzkow (2012).

In this article, we revisit the gains from variety made available by e-commerce for a commonly purchased good, shoes. We begin by collecting new data containing millions of geographically disaggregated footwear sales, daily inventory, and all product reviews from a large online retailer. We present descriptive evidence that heterogeneity in consumer tastes across markets is substantial. As a result, we would expect that local retailers respond to these differences in local demand across locations and show supporting evidence of this using supplementary brick-and-mortar assortment data. We then employ a structural model of demand to obtain empirical estimates of the value of increased variety due to online retail. Our results show that omitting the role of local tastes and retailer responses leads to estimates that significantly overstate the gains from online variety. In our application, we find this overstatement to be over 75%.

Our results have two major policy implications. First, the disproportionate impact of variety on welfare found in the previous literature suggests that antitrust enforcers and policy makers should weigh potential changes in variety more than potential changes in price. We find that although the gains from access to variety are still meaningful, they are not significantly greater than the estimated gains from lower prices due to e-commerce. For example, if online retail has led to a 5% decrease in prices, we find that lower prices would account for just over 50% of the consumer welfare gains derived from online retail. That is, we estimate the variety effect to be about equal in size to the price effect whereas the previous literature estimates the variety effect to be seven to ten times larger than the price effect. A second, related implication is that such a large increase in welfare from variety suggests consumers could endure a significant negative income shock and still be as well off as before online retail. In other words, the compensating variation of the additional variety is negative and large in magnitude. Thus, an implication of Brynjolfsson, Hu, and Smith (2003) is that online retail has led to a massive decline in the price index for books. If this effect holds generally across online retail sectors, this may suggest the consumer price index (CPI) has also seen a rapid decline. However, our results suggest this is unlikely to be the case.

To obtain these estimates, we must confront empirical challenges that arise from infrequently purchased products. Although demand estimation techniques, such as Berry (1994) and Berry, Levinsohn, and Pakes (1995), have been very successful in producing sensible estimates with aggregated data (across geographic markets, time, and/or products), it is typically problematic to apply these techniques to very granular data. With high-frequency sales data, such as our own, the number of available options typically rises as fast (or faster) than the number of purchases. As a result, for any given product, we may observe few, if any, sales and this will be the case for thousands of products. This means if we take the observed market shares as proxies for the true underlying choice probabilities, they will likely be observed with error. Because the true underlying choice probabilities are unknown, this small sample issue is readily apparent to the researcher only when no sale, opposed to few sales, is observed.

Two approaches are commonly used to force data that suffers from small sample sizes into existing estimation techniques. However, we show both are unsatisfactory when the goal is to estimate the demand for narrowly defined products across narrowly defined markets. The first is to aggregate data over markets or products until observations with zero sales disappear. We show that aggregation exactly smooths over the heterogeneity of interest. The second approach is to ignore the issue and simply omit observations with zero sales from the analysis (such as, focusing on just popular products). Omitting observations without sales assumes that there is no demand for these products and is problematic for two reasons. First, it creates a selection bias in the demand estimates (Berry, Linton, and Pakes 2004, Gandhi, Lu, and Shi 2013, Gandhi, Lu, and Shi 2014), which tends to result in estimating consumers as too price inelastic. Second, for our setting this is particularly problematic because if uncorrected, we would overstate the degree of heterogeneity across markets (Ellison and Glaeser 1997), leading us to understate the gains from online variety. For example, if we observe only one shoe sale in a particular market, it would suggest there are no gains from additional variety because only one particular product is desired.

We address these concerns by augmenting the nested logit model with across-market random effects, which allows us to decompose product-level unobservables into a mean national-product-level fixed effect and local market deviations. With a parametric assumption on the deviations, we show that aggregating to the national level leads to an estimating equation that involves only national-product-level market shares, local-nest market shares, and the distribution of local deviations. Importantly, the estimator does not rely on the local market shares of individual products, the vast majority of which are zero at each location, but allows us to retain information about the distribution of local demand heterogeneity. Our results lie between the extremes of dropping all of the zeros and adjusting all of the zeros by the same amount (Gandhi, Lu, and Shi 2014). As a result, for example, the unsold boot in Alaska is treated differently than the unsold sandal in Alaska. Additionally, the method addresses the well-known econometric challenge that logit-style demand models tend to overstate welfare gains under large changes in the choice set as each new product introduces a new dimension of unobserved consumer heterogeneity (Ackerberg and Rysman 2005). We compare our estimates to existing approaches and show it performs well.

To identify the distribution of the random effects, we appeal to what is commonly viewed as a data problem – zero shares – as the source of identification. Note that our random effects model does not by itself predict or explain zero market shares, rather, local zeros are rationalized by modeling the finite multinomial that generates purchases at the local level. Our across-market heterogeneity parameters are then chosen to match the model's predictions to a set of micro moments (Petrin 2002, Berry, Levinsohn, and Pakes 2004) that are based on the observed fraction of local zeros. For example, if a product experiences a large number sales but also many locations with no sale, this would suggest that the demand for this product is concentrated in select markets and the preference heterogeneity for this product is high. The design of our micro moments makes use of millions of data points, allowing us to identify a relatively large number of parameters.

Estimating our model allows us to quantify the magnitude of across-market demand heterogeneity and we find that demand varies greatly across markets. For example, a one standard deviation increase in the local demand shock of a sneaker is equivalent to a decline in price of $34 for men and $62 for women. As a result, in our data, we find that the estimates of consumer gains from online variety are about 45% smaller when accounting for local demand heterogeneity and the retailer responses to it. In other words, if local stores mostly satisfy local demand, then the incremental value of online markets is much smaller because the average consumer already has access to most of the products he or she wants to purchase. Taken together, our results suggest that the "long tail" pattern observed in online retail is simply a consequence of demand aggregation, where separate local demand curves are aggregated over geographies.

The rest of the article is organized as follows. Section 2 discusses our data and presents descriptive evidence of across-market demand heterogeneity and supply side responses. Section 3 introduces the model. Section 4 describes the estimation procedure. Results and counterfactuals are in Sections 5 and 6, respectively. Section 7 discusses the robustness of our findings, and the conclusion follows.

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. In 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). Next, we provide evidence of the localization of assortments using the brick-and-mortar assortment data (Section 2.2) and then demand-side across-market heterogeneity using the online retail sales data (Section 2.3). Finally, we document the small sample problem in the sales data – in particular, the zeros problem – and show simple aggregation cannot satisfactorily 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 information about the shoe, including model and style information. The transaction identifier allows us to see if an order contains more than a single pair of shoes, but we observe no additional information about the customer. Finally, we download a picture of each shoe and image process it to create color covariates.

We observe over 13.5 million shoe transactions during the collection period, with two-thirds of transactions being women's shoes. The price of shoes varies substantially both across gender and within gender – for example, dress shoes tend to be more expensive than sneakers. 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.

The sales data is merged with product review and inventory data. The review data contain a time series of reviews and ratings for each shoe. 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 the average ratings for comfort, look, and overall appeal, where 1 is the lowest rating and 5 is the highest rating. These ratings are heavily skewed towards favorable ratings. We treat these variables as time varying features of the product that capture information available to the consumer at the time of purchase.

In the inventory data, we track daily inventory for every shoe.[2] 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 (Conlon and Mortimer 2013). Although the inventory data is size specific, the sales data does not include size. We concede that this, in general, will cause us to understate the gains from online variety because consumers with unusual foot sizes may greatly benefit from online shopping if traditional retailers do not typically stock unusual sizes.[3] The average daily assortment size is over 50,000 products, but, over the span of data collection, over 100,000 varieties of shoes were offered for sale. This suggests that there is significant turnover in the choice set, with some products being offered over the entire sample and others appearing for brief periods of time.

Brick-And-Mortar Data

In addition to the online retail sales data, we collect a snapshot of shoe availability from Macy's and Payless ShoeSource during August and September of 2014. Although these chains have different business models and cater to different types of consumers, we find and highlight patterns in both of their assortment decisions that are consistent with local customization.

For each retailer, we began by collecting all of the shoe SKUs the 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 at each location. The firms' websites do not list how many shoes are in stock, just whether a shoe is in stock or not. In addition to a shoe identifier, which is unique to each chain, we are also able to obtain the brand, category, and color of the shoe.[4] As each query was for a specific shoe size, we then aggregate across all sizes to have a measure of product availability consistent with our product definition. Aggregating over sizes also lessens the possibility that our analysis is skewed by particular sizes being temporarily out-of-stock. We cannot merge this brick-and-mortar data with our online sales data as the collection periods do not overlap and the firms utilize different product identifiers. Payless also offers many exclusive varieties. However, we can use the data to examine local assortments.

[2]Initially this data was not collected daily, but for the last seven months of data collection, shoe inventory was tracked daily. Prior to daily collection, inventory was imputed by assuming a product was in stock between its first and last stock or sale dates.

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

[4]We did not scrape webpages but rather downloaded the information by targeted server queries. Hence, the information we are able to obtain is limited.

Table 1 presents summary information on the assortments of 649 Macy's locations and 3,141 Payless' locations. In September 2014, we observe 13,914 different styles available at Macys.com, of which about 42% are online exclusives. At Payless.com, we observe 1,430 distinct styles, with about 19% being online exclusives. Average in-store assortment sizes are 871.7 and 513.0 for Macy's and Payless, respectively. There is a much greater variance in Macy's store sizes.[5] Unsurprisingly, we find that the stores with larger assortments tend to be located around larger population centers.

2.2 Localization of Brick-And-Mortar Retailers

Brick-and-mortar retailers are known for offering different product assortments across their network of stores. Figure 1 shows histograms of shoe availability across store locations for Macy's and Payless. The Macy's distribution is heavily skewed toward the left, meaning most shoes are available at only a few stores. The Payless distribution is bimodal, with many shoes available at only a few stores and others offered at all locations. For large national retailers, there are trade offs to localizing assortments. On the one hand, catering to local demand may greatly increase revenues, but on the other hand, there are cost advantages from economies of scale through standardization. Available evidence suggests the former may outweigh the latter. For example, in recent years, Macy's has made a concerted effort to better localize its product assortments through a program called "My Macy's:"

> "We continued to refine and improve the My Macy's process for localizing merchandise assortments by store location.... 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."[6]

Of course, a firm's words may differ from their actions and although we see large differences in assortments across stores, this may be due to variation in store sizes. To calculate a measure of assortment similarity, we take the network of stores within a particular chain and create all possible links between stores. Then for each pair of stores with assortment sets (A, B), we calculate

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

[5] According to a Macy's investor file, the standard deviation in size across Macy's locations is 149,000 square feet, where the 5th-percentile store is 47,000 square feet and the 95th-percentile is 325,000 square feet.

[6] https://www.macysinc.com/macys/m.o.m.-strategies/default.aspx

The numerator is the number of products the two stores have in common and the denominator is the minimum of the two stores' assortment sizes. This gives us a measure that is bounded between zero and one, with zero meaning no assortment overlap and one implying complete overlap, adjusting for differences in store size.[7] To further isolate differences in variety from differences in assortment size, we directly compare only locations of similar size. Figure 2 plots Lowess fitted values of this exercise for Macy's and Payless as a function of distance between stores A and B. We can see that the assortment overlap has a decreasing relationship with distance suggesting these retailers localize their product assortments. Additionally, Macy's stocks more sandals (up to 10%), as a percentage of local assortment, in warm weather locations and more boots (up to 4%) in cold weather locations. There is also significant heterogeneity in brands across locations as the average Macy's store stocks less than half of the 160 brands in the data.

We acknowledge there may be some supply-side factors that affect the differences we observe in assortments. For example, as distance approaches zero, assortment similarity does not converge to one. This may reflect a strategy to increase variety within a geographic area when individual stores face limited floor space,[8] in addition to some locations where retailers maintain separate men's and women's stores. Further, manufacturers may influence the availability of products across space, as Thomas (2011) shows for large manufacturers across countries. However, we do not believe there exists a substantial difference in relative costs across products that could lead to this geographic pattern because the vast majority of products are imported.

2.3 Across-Market Demand Heterogeneity in Online Data

In our online retail data, the observed prices, product characteristics, and choice sets are the same for all markets, suggesting differences in observed local market shares can be rationalized only by differences in local demand (or by sampling, which we address shortly). In Table 2, we present the local and national share of revenue generated by the top 500 products ranked within a local market. For example, suppose we defined a market as a combined statistical area plus the remaining parts of the states (CSA+state).[9] At the CSA+state-month level, we observe 213 local markets over 14

[7]If, for example, there were no across-market demand heterogeneity and stores stocked products in order of most to least popular, our measure would find complete overlap. Other measures, such as cosine similarity (Hwang, Bronnenberg, and Thomadsen 2010), would differ from complete overlap, but this would be entirely driven by differences in store size.

[8]In our analysis to follow, we allow for this possibility by attempting to proxy the number of products available in each local market, rather than at a particular store.

[9]There are 165 CSAs, which are composed of adjacent metropolitan and micropolitan statistical areas. We then define states as the portion of a state not contained in a CSA. This adds an additional 48 markets. All of Rhode Island and

time periods. On average, the top 500 products at this disaggregated level make up 67.05% of local revenue. If we take the same 500 products and calculate their national-level revenue share, on average, they make up only 7.19% of national revenue.

If demand were homogeneous across markets, we would expect the share of revenue accruing to these products to be the same locally and nationally. The extent to which they differ provides evidence that people in different locations demand different products. For most definitions of the local market, there are large differences between the local revenue share and the national revenue share. This suggests that the commonality of popular products is quite small across markets.[10]

We formally test for across-market demand heterogeneity, controlling for local sample size, using multinomial tests comparing local market shares ($s_{\ell j}$) to national market shares (s_j). Define $s = \{s_j\}_{j=1}^J$ and $s_\ell = \{s_{\ell j}\}_{j=1}^J$, then the null hypothesis is $H_0 : s = s_\ell$. The last column in Table 2 presents the rejection rates for various levels of aggregation. We can see that these tests are overwhelmingly rejected at all levels of aggregation. However, the tests reveal effects coming from both zeros and aggregation. At more disaggregated levels, zeros become more prevalent, reducing the power of the multinomial tests (e.g. zip5 rejection rate < zip3 rejection rate). At the other end of the spectrum, aggregating up to Census Regions greatly obscures heterogeneity across markets leading to a slight reduction in rejection rates when compared to the state level (94% vs. 92%).

Some differences in demand across markets occur for obvious reasons. Take our earlier example of boots versus sandals. Figure 3 plots average annual state temperature against state-category market shares of 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. This also suggests that consumers do not shop online just for products that are not available in traditional brick-and-mortar stores. For example, sandals make up a larger share (and boots a smaller share) of shoe revenue in Florida and Hawaii than in colder weather states.

Other differences in demand across markets occur for less obvious reasons. In Figure 4, we map the consumption pattern of a popular brand by national revenue. Local revenue share at the 3-digit zip code level is mapped for the eastern United States. Although this brand is popular when measured by national sales, we can see a clear preference for this brand in the Northeast. In Florida this brand makes up less than 2.5% of sales, whereas in parts of New York, New Jersey,

New Jersey are contained in a CSA.

[10] A small cutoff (500, or 1% of products) was chosen to single out popular products and limit the impact of sampling. We also conducted this analysis with cutoffs ranging from one to over 50,000 and find intuitive results. The percent differences decreases as the cutoff increases between 3,000-5,000.

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

2.4 Aggregation and the Zeros Problem

The vast majority of products have local market shares equal to zero. Table 3 shows the severity of the zeros problem in our data. At fine levels of geography, such as defining a market at the 5-digit zip code-month level, 99.96% of products have zero sales. Although simple aggregation over geographies does alleviate the zeros problem, what is astonishing is that even at highly aggregated levels, such as state-month, 85.25% of products have zero sales. Furthermore, Table 2 shows for high levels of aggregation, the heterogeneity we are interested in exploring is effectively smoothed over, as the revenue share comparison of the top 500 products becomes increasingly similar. Further, aggregation over product space produces equally poor results (Table 4).

3 Model of Consumer Demand

Within each location, demand follows a standard nested logit demand framework. 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 any product. Let c index a set of nests that partition the set of products $J \cup \{0\}$ with the outside good belonging to its own nest.

To ease notation, the time script t has been suppressed. For a product j, the utility of a consumer $i \in I_\ell$ in location $\ell \in L$ is given by

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

where $\delta_{\ell j}$ is the mean utility of product j at location ℓ, $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution and, for consumer i, ζ_{ic} is common to all products in the same category and has a distribution that depends on the nesting parameter λ, $0 \leq \lambda < 1$. Cardell (1997) shows that there exists a distribution of ζ_{ic} such that $\zeta_{ic} + (1 - \lambda)\varepsilon_{i\ell j}$ has a generalized extreme value (GEV) distribution, leading to the frequently used nested logit demand model. The parameter λ determines the within category correlation of utilities. When $\lambda \to 1$ consumers will substitute only to products within the same nest and when $\lambda = 0$ the model collapses to simple logit.

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

$$\delta_{\ell j} = x_j\beta - \alpha p_j + \xi_{\ell j},$$

where x_j is a vector of product j's characteristics, p_j is product j's price, and $\xi_{\ell j}$ is a location-specific unobserved product quality. Observable characteristics do not differ across locations and we assume preferences over observable characteristics are constant across locations. Therefore, demand across locations differs only by the location-specific unobserved product qualities.

Integrating over the GEV 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 outside good has utility normalized to zero, i.e. $\delta_{\ell 0} = 0$, $\forall \ell \in L$. The choice probabilities have the following analytic expression:

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

where $\pi_{\ell c}$ is the location-specific probability of choosing any product in category c and $\pi_{\ell j/c}$ is the location-specific probability of choosing j conditional on choosing c.

As shown in Berry (1994), the choice probabilities can be inverted revealing a linear equation to be estimated:

$$\log(\pi_{\ell j}) - \log(\pi_{\ell 0}) = x_j\beta - \alpha p_j + \lambda \log(\pi_{\ell j|c}) + \xi_{\ell j}, \tag{2}$$

This allows linear instrumental variables techniques to be applied to control for price endogeneity. In the standard estimation of this model with location-level data, the $\xi_{\ell j}$'s are nonparametrically identified as the residuals of the estimation. The maintained assumption is that the size of each local market ℓ is sufficiently large so that $\pi_{\ell j}$ and $\pi_{\ell j|c}$ are observed without error, for all products j. With highly aggregated markets or a small number of products this market size assumption may be reasonable. However, in high-frequency, highly disaggregated sales data we often find that the number of available options is large relative to the number of observed purchases. This suggests the existence of a small sample problem that breaks the link between observed market shares and true underlying consumer choice probabilities. Hence, estimates relying on observed market shares as proxies for consumer choice probabilities will result in biased estimates. If market shares were merely small, this issue may go unnoticed. However, it is common in data to find a large number of products with zero market share. This is the extreme manifestation of the small sample problem and the most visible to researchers because it creates a mechanical problem of having to take the log of zero. When confronted with small sample sizes the precise location-specific $\xi_{\ell j}$'s cannot be identified.

To make further progress, we place additional structure on the location-specific mean utilities, $\delta_{\ell j}$, by assuming that the location-specific unobserved qualities, $\xi_{\ell j}$, are additively separable in two components. An average term that is constant across locations, ξ_j, and a location-specific deviation, $\eta_{\ell j}$, which we assume is drawn independently from a normal distribution, $N(0, \sigma_j^2)$. That is, we decompose the unobserved product quality into a mean aggregate-product-level fixed effect and a local-product-level deviation. Rearranging terms we have,

$$\delta_{\ell j} = \underbrace{x_j\beta - \alpha p_j + \xi_j}_{\delta_j} + \eta_{\ell j}.$$

Here δ_j is the mean utility of product j for the (national) population of consumers. In the following subsection, we show that national mean utilities can be written as a function of aggregate (national) market shares, local-nest market shares, and the distribution local deviations. In particular, by aggregating over locations, the local market shares of individual products drop out of our derivation. This allows us to avoid the small sample issue by retaining information on across-market demand heterogeneity through the distribution of local deviations instead of trying to identify exact location-product realizations.

Heterogeneity in preferences among consumers in our model comes from two sources, an "across-market" effect, $\eta_{\ell j}$ and a "within-market" effect, $\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}$. If $\eta_{\ell j} = 0$ for all $\ell \in L, j \in J$, the model reduces to a standard nested logit model, where there is no distinction between local and national preferences. As σ_j increases in the model, demand becomes more dispersed across markets and less dispersed within markets. This suggests higher variances coincide with situations where targeting by local retailers is both easier and more valuable.

3.1 Aggregation Theory and the Market Share Inversion

To address the sampling concern, we would like to aggregate over locations, while retaining information about the heterogeneity in local demand. The structure of our model will allow us to do this. By appealing to the law of large numbers in locations, Proposition 1, below, states that the weighted summation of local demand implied by any set of ηs is a convergent series. The proof is in Appendix A.

Proposition 1. *For each product $j \in J$, applying the law of large numbers (for weighted sums with population weights ω_ℓ^L) in L and integrating out over η gives*

$$\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell j}(\eta_\ell; \delta, \lambda) - \pi_j \rightarrow_{a.s.} 0 \tag{3}$$

This implies, with a large number of locations, we can recover the weighted summation term without knowing the exact realizations of η. Therefore, aggregated choice probabilities depend only on the variance of the across-market heterogeneity and 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,

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

Once δ is obtained, standard instrumental variables methods, instrumenting for price, can be used.

The inversion for our random effects model with L locations takes the form (see Appendix A for derivation),

$$\delta_j(\pi; \lambda, \sigma) = \delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \log\left(\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right). \tag{4}$$

Equation 4 relates δ_j to product j's aggregated share, π_j, local population shares, ω_ℓ^L, local outside good and category shares, $\pi_{\ell 0}$ and $\pi_{\ell c}$, and the random effect, $\eta_{\ell j}$. Additionally, note that this inversion reduces to the inversion found in Berry (1994) when $\eta_{\ell j} = 0, \forall \ell \in L, j \in J$. [11] However, because $\eta_{\ell j}$ is an unknown random variable, unlike Berry (1994), we cannot simply recover mean

[11]Suppose $\eta_{\ell j} = 0, \forall \ell \in L, j \in J$, then $\pi_{\ell 0} = \pi_0$ and $\pi_{\ell c} = \pi_c$, and

$$\delta_j = (1-\lambda)\left(\log \pi_j - \log\left(\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right) = (1-\lambda)\left(\log \pi_j - \log\left(\sum_{\ell=1}^{L} \omega_\ell^L \pi_c^{-\frac{\lambda}{1-\lambda}} \pi_0^{\frac{1}{1-\lambda}}\right)\right)$$

$$= (1-\lambda)\log \pi_j + \lambda \log \pi_c - \log \pi_0 = \log \pi_j - \log \pi_0 - \lambda \log\left(\frac{\pi_j}{\pi_c}\right) = \log \pi_j - \log \pi_0 - \lambda \log \pi_{j|c}$$

utilities from observables.

To integrate out over the $\eta_{\ell j}$s, first note that the LLN applied in Proposition 1 implies

$$\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\} - \sum_{\ell=1}^{L} E\left[\omega_\ell^L \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] \rightarrow_{a.s.} 0.$$

The complexity of this expectation is highlighted when we apply the Law of Iterated Expectations,

$$E\left[\omega_\ell^L \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] = E\left[\omega_\ell^L \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\} \middle| \pi_{\ell c}, \pi_{\ell 0}\right]\right].$$

The conditional expectation depends not only on the local mean utilities of all other products, but also the conditioning variable is the sum of lognormal random variables, which does not have a closed form expression for its distribution. We appeal to the assumption that there are a large number of products for each category in order to make further progress, as stated in Proposition 2 below.

Proposition 2. *Suppose the law of large numbers applies, i.e.* $\frac{1}{J_c}\sum_{j\in c} \exp\left\{(\delta_j + \eta_{\ell j})/(1-\lambda)\right\}$ *converges in distribution to a constant for each c, then*

$$E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\} \middle| \pi_{\ell c}, \pi_{\ell 0}\right] \rightarrow_d E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right], \quad as \quad J \rightarrow \infty.$$

The difficulty in Proposition 2 comes from the fact that the conditioning arguments are not independent and identically distributed. If there is convergence, it should be to the expectation; however, convergence is not obvious. Proposition 2 requires that products are added at a rate proportional to category shares for each category. The benefit of appealing to a large number of products for each category is that it allows for approximating the conditional expectation with the unconditional, which is simple to compute using the moment generating function of the normal distribution,[12] $E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] = \exp\left\{\frac{1}{2}\frac{\sigma_j^2}{(1-\lambda)^2}\right\}$. Intuitively, when more products are added to a market the sum of random demand shocks is less informative about any individual shock. In Monte Carlo exercises, using the unconditional expectation to approximate the conditional expectation performs well.[13]

[12] The moment generating function of a normal distribution with mean μ and variance σ^2 is,

$$E\left[\exp\{xt\}\right] = M_x(t) = \exp\left\{\mu t + \frac{1}{2}\sigma_j^2 t^2\right\}.$$

[13] See the appendix for details.

Finally, although small sample sizes make the observed local market shares unreasonable estimates of the true underlying choice probabilities for individual products, we assume the national choice probabilities, π_j, the local choice probabilities of the outside good, $\pi_{\ell 0}$, and the local category choice probabilities, $\pi_{\ell c}$, are well estimated and strictly positive in the data. This is reasonable if the size of the population is large relative to the number of categories. With these assumptions and given any (σ, λ), we can then recover national mean utilities as a function of observables $(\pi_j, \pi_{\ell c}, \pi_{\ell 0})$,

$$\delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \frac{1}{2}\frac{\sigma_j^2}{(1-\lambda)^2} - \log\left(\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\right)\right). \tag{5}$$

Proposition 2 has two important implications. First, it allows for the introduction of random coefficients, while still allowing δ_j to be recovered point-wise. This contrasts with Berry, Levinsohn, and Pakes (1995), henceforth BLP, which requires a $J \times J$ system of equations to be simultaneously solved for each location ℓ. Our approach greatly reduces the computational burden of the problem, especially in situations with a large number of products and locations. BLP introduces random coefficients through the interaction of product characteristics and consumer demographics. However, to the extent that observable demographics fail to capture differences across markets, the degree of across-market heterogeneity will be understated and the estimated gains from online variety will be overstated. Indeed, it is likely that observable demographics will not fully capture differences in tastes across locations (Bronnenberg, Dhar, and Dube 2009, Bronnenberg, Dube, and Gentzkow 2012).

The standard BLP approach would normally address this by using local market shares to estimate the preference parameters and then use the local-level residuals to form an estimate of η, much like the standard nested logit model. However, this also would suffer from the sampling problem noted previously. For large data sets, the sampling problem is likely to be severe as the number of products relative to purchases tends to be high. Our approach addresses the problems that arise from sampling error, while retaining information about across-market demand heterogeneity.

An alternative approach to bypass the issue of market-level zeros is to estimate the individual demand of products via maximum likelihood. This could be appropriate in our setting as we have individual-level purchases (but no individual-level covariates). The likelihood would be comprised of modeling individual purchase decisions, i.e., the probability that i chooses $j \in J \cup \{0\}$.

Here, market-level zeros are not an issue per se, as products not purchased are simply dropped from the likelihood. However, as noted in Chintagunta and Dube (2005), there is a significant computational burden in estimating such a model, while accounting for price endogeneity. To make progress, they augment the maximum likelihood estimator by imposing the market share inversion of BLP. Goolsbee and Petrin (2004) also utilizes the uniqueness of the inversion in estimating such a model. Of course, this inversion does not exist when shares are observed to be zero. Thus, additional assumptions are required, such as including as many covariates as possible and assuming this adequately removes the price endogeneity or assuming a particular form of the endogeneity. Our method allows for both price endogeneity and local-level zeros in a computationally tractable way.

Second, our approach relates to the work of Ackerberg and Rysman (2005), who provide a method for disciplining the role of unobserved individual heterogeneity (the εs) in demand estimation. This is particularly important in our setting as logit-style demand models tend to overstate the value of adding products (as a result of taking the max over many draws of ε). To see the relationship, define

$$R(\sigma_j) = \exp\left\{\frac{1}{2}\frac{\sigma_j^2}{(1-\lambda)^2}\right\}.$$

As $R(\sigma_j)$ is not indexed by ℓ, the share equation can be rearranged to yield

$$\pi_j = R(\sigma_j)\exp\left\{\frac{\delta_j}{1-\lambda}\right\}\sum_{\ell=1}^{L}\omega_\ell^L \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}.$$

Expanding this equation, we obtain[14]

$$\pi_j = \frac{\left(\sum_{j'\in c} R(\sigma_{j'})\exp\{\delta_{j'}/(1-\lambda)\}\right)^{1-\lambda}}{1+\sum_{c'\in C}\left(\sum_{j'\in c'} R(\sigma_{j'})\exp\{\delta_{j'}/(1-\lambda)\}\right)^{1-\lambda}}\cdot\frac{R(\sigma_j)\exp\{\delta_j/(1-\lambda)\}}{\sum_{j'\in c} R(\sigma_{j'})\exp\{\delta_{j'}/(1-\lambda)\}}. \tag{6}$$

That is, at the national level, the local random effects can be summarized as a function of the

[14]To see this, note that

$$R(\sigma_j)\exp\left\{\frac{\delta_j}{1-\lambda}\right\} = \exp\left\{\frac{\delta_j + (1-\lambda)\log R(\sigma_j)}{1-\lambda}\right\}.$$

Define $\tilde{\delta}_j = \delta_j + (1-\lambda)\log R(\sigma_j)$. Plugging this into the expanded nested logit share equation gives

$$\pi_j = \frac{\left(\sum_{j\in c}\exp\{\tilde{\delta}_j/(1-\lambda)\}\right)^{1-\lambda}}{1+\sum_{c'\in C}\left(\sum_{j'\in c'}\exp\{\tilde{\delta}_{j'}/(1-\lambda)\}\right)^{1-\lambda}}\cdot\frac{\exp\{\tilde{\delta}_j/(1-\lambda)\}}{\sum_{j'\in c}\exp\{\tilde{\delta}_{j'}/(1-\lambda)\}}.$$

Finally, substituting back in for $\tilde{\delta}_j = \delta_j + (1-\lambda)\log R(\sigma_j)$ gives us Equation 6.

variances. Equation 6 has a striking similarity to the nested logit formulation in Ackerberg and Rysman (2005); however, the interpretation and implementation differs from our approach. In Ackerberg and Rysman (2005), the penalty must be specified by the econometrician; they use an assumption on the number of retail outlets per product. Our penalty term is a summary statistic of across-market demand heterogeneity. This allows us to motivate and identify the penalty term using observable micro data on across-market demand heterogeneity. It is important to note that, although both methods are similar at the aggregate level, modeling disaggregated (local) demand was not an objective of Ackerberg and Rysman (2005). Indeed, estimating local demand using Ackerberg and Rysman (2005) would result in the same small sample problem previously noted. The advantage of our approach is that it generates crowding at the national level, but also allows for estimating the distribution of local preferences.

4 Estimation

4.1 Micro Moments

To identify the random effects, we need additional moments that capture the differing degrees of across-market heterogeneity among products. In our model, σ alters the degree of local concentration in demand. A higher σ creates greater extremes in location-specific draws suggesting local demand that is more concentrated in the subset of products that have very high draws of η. This pulls away sales from all other products in that local market. Thus, for most products in a market, the probability of not observing a sale will increase as the demand becomes concentrated in the high draw products. Because the fraction of local markets with very high draws for a particular product will be small, overall, the fraction of markets where no sales of that product occur will be increasing in σ.[15]

Although we have emphasized that zero sales are normally problematic when estimating demand, the above suggests that we can appeal to them as the source of identification of across-market demand heterogeneity. Let $P0_{\ell j}(\sigma;\delta,\lambda)$ be the probability that a product j has zero sales, given N_ℓ, the population of consumers at location ℓ. We then define

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

[15]We show this graphically with our estimated model in the robustness section. There is a monotonic relationship between the proportion of zero sales and the magnitude of across-market demand heterogeneity.

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

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

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.

$$mm_j(\sigma; \delta, \lambda) = \left(P0_j(\sigma; \delta, \lambda) - \widehat{P0}_j\right).$$

It is important to point out that $P0$ is just one such micro moment that can be used to estimate across-market demand heterogeneity. Other moments include $P1, P2$, etc., as well as the 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 fraction of locations may experience zero sales for a given product.[16]

4.2 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 generalized method of moments (GMM). We start with the implementation of the micro moments. Note that local-level mean utilities can be written as

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

where where each location, ℓ, receives an i.i.d. draw, $\bar{\eta}_{\ell j}$, from a standard normal distribution. Given the assumptions on the individual-level unobservable ($\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}$ is GEV), the location-product-level choice probabilities take the standard nested logit closed form expression, i.e.

$$\hat{\pi}_{\ell j}(\sigma; \delta, \lambda) = \frac{\left(\sum_{j' \in c} \exp\{(\delta_j + \sigma_j \bar{\eta}_{\ell j})/(1-\lambda)\}\right)^{1-\lambda}}{1 + \sum_{c' \in C}\left(\sum_{j' \in c'} \exp\{(\delta_{j'} + \sigma_{j'} \bar{\eta}_{\ell j'})/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{\exp\{(\delta_j + \sigma_j \bar{\eta}_{\ell j})/(1-\lambda)\}}{\sum_{j' \in c} \exp\{(\delta_{j'} + \sigma_{j'} \bar{\eta}_{\ell j'})/(1-\lambda)\}}, \quad (7)$$

[16]The logit structure implies $P0$ is no longer valid when assuming large N for all locations because each product will then have positive local share.

for any candidate value of σ and λ. The probability a product is observed to have zero sales at location ℓ is then

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

i.e. the probability all N_ℓ consumers at location ℓ choose not to purchase good j. We then average over locations and match it to the fraction of locations observing zero sales of j. This approach is computationally fast and avoids the problems posed by simulating individual purchase decisions, which is useful for big data sets that contain a large number of locations and products.

With a candidate solution for σ and λ, the structure we have placed on the ηs allows us to integrate them out according to Equation 5 and recover national mean utilities

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

Hence, we obtain a linear equation to estimate where instrumental variable methods can be used to control for price endogeneity.

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 5,

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

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.[17] The same instrument can be employed here, because $z_{j|c}$ is correlated with $\frac{\partial \delta_j(\pi,\lambda,\sigma)}{\partial \lambda}$ 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.

Let Z be the usual matrix of nested logit instruments that identify β, α, λ and denote the set of moments, $m = E[Z'\xi]$. Stacking the moments and micro moments where $\theta = (\sigma, \lambda, \beta, \alpha)$, we have

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

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

and the GMM criterion is $G(\theta;\cdot)^T W G(\theta;\cdot)$, with weighting matrix W. In the first stage, we take $W^0 = \left(G(\theta^{(0)};\cdot)G(\theta^{(0)};\cdot)^T\right)^{-1}$ for an initial value, $\theta^{(0)}$.[18] Then using the solution from the first stage, $\widehat{\theta}^{(1)}$, we use $\widehat{W} = \left(G(\widehat{\theta}^{(1)};\cdot)G(\widehat{\theta}^{(1)};\cdot)^T\right)^{-1}$ in the second stage. Our final estimates are $\widehat{\theta}^{(2)}$.

5 Results

In this section, we discuss our demand estimates and the fit of the model. We restrict our attention to adult shoes and estimate the demand for men's and women's shoes separately. The size of the choice sets average about 13,250 and 27,500 products for men and women, respectively. We define our time horizons to be at the monthly level and our geographic locations to be composed of 213 local markets (165 CSAs plus 48 states).[19] We choose CSA+state, compared to just CSA, because a large percentage of observed sales occur outside CSAs. For example, if we pursued the CSA market definition, we would drop all of the sales to consumers in Alaska. Our market sizes (N_ℓ) are equal to the local population for men and women from the American Community Survey,[20] with the interpretation being that each consumer makes a single purchasing decision every month.[21]

Included in x are product ratings for comfort, look, and overall appeal and fixed effects for color, brand,[22] and time. The product ratings are time varying and reflect what the consumer would observe at the time of purchase. We instrument for both price and the within group share using the typical BLP-style instruments. Included are the number of available styles (color combinations) for a particular shoe model, and the sum and average of within-category competitor characteristics. That is, let B denote the set of brands and c_b denote the set of shoes manufactured by brand $b \in B$ in category $c \in C$. For each time period, our additional instruments are

$$\sum_{j' \neq j \in c_b} x_{j'}, \quad \frac{1}{\# c_b} \sum_{j' \neq j \in c_b} x_{j'}.$$

These will aid us in identifying the price coefficient, α, and the nesting parameter, λ.

[18] We repeat the estimation for a set of randomly drawn θ^0. We also take $W^0 = I$, but find specifying an initial weighting matrix decreases the computational time.

[19] We find at finer levels of geography, such as zip code, the nearly 100% local zeros cause the micro moments to lose identifying power.

[20] More specifically, zip code-level gender-specific population data come from the American Community Survey. Zip codes are then mapped to CSA's and N_ℓ is then the male or female population summed over zip codes contained in that CSA. Zip codes outside of CSAs are aggregated to the state level.

[21] According to an American Apparel & Footwear Association report, the average American purchased 7.5 pairs of shoes in 2013.

[22] More specifically, we create fixed effects for brands that average at least 50 sales per month and group the remaining smaller brands. This results in 213 brand fixed effects for men and 331 for women. We use the within transformation along the brand dimension.

In principle, with our modeling assumptions and a large number of product-location observations, we could estimate J random effects, i.e. a σ_j for each individual product. However, the large observed choice set would create a significant computational burden in estimating individual product-level heterogeneity parameters. Thus, for empirical tractability, we parameterize σ as a category-summer and category-winter random effect for boots and sandals, and as a category random effect for all other categories,[23]

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

Thus $mm(\cdot)$ contains $C + 2$ moments.[24] Overall, the estimation of the augmented nested logit model involves identifying up to twelve random coefficients (across-market parameters), the nesting parameter (λ), as well as the the remaining 251-371 mean utility parameters (α, β) using 1.8-3.9 million product-market-level observations.[25]

We compare the estimates of our approach with a number of alternative models that are commonly used to estimate demand. For ease of exposition, we define these approaches now:

Local RE	Location-product-level random effect model (our approach)
Local NL	Traditional nested logit model at the local level
National NL	Traditional nested logit model at the aggregated (national) level

We estimate the Local NL model for two sets of data. The first treats observed shares as true shares and drops all of the zeros. We call these unadjusted shares. We present these results for comparison because this is the standard approach when confronted with zeros. The second data set adjusts aggregate zeros using the correction proposed by Gandhi, Lu, and Shi (2014), which we call adjusted shares. The purpose of the adjustment in Gandhi, Lu, and Shi (2014) is not only to bring the zeros off the bound, but to do so in an "optimal" fashion. The procedure is based on a Laplace transformation of the observed shares, with additional steps to minimize the asymptotic bias between the adjusted shares and the true conditional choice probabilities. Their approach is useful in a variety of settings, and we apply it to the relatively few zeros in the aggregate

[23]This is motivated by observations in our sales and inventory data. The fraction of sales and the fraction of the choice set made up by sandals spikes in the summer and troughs in the winter. The reverse is observed in boots, whereas all other categories remain relatively stable over the course of the year.

[24]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.

[25]Estimation takes about a day with an Intel Xeon E5-1650 processor running at 3.5GHz using analytic gradients and the Knitro solver.

data. These zeros are due to products entering or exiting the choice set within the month. For example, a sneaker may come into stock with only a day left in the month, and consequently, we do not observe any sales for that period. However, we find that this asymptotic correction has little impact on the local data because of the very large fraction of zeros (see results for Local NL-Adjusted). Importantly, for our local-level analysis, the correction adjusts all local zeros by the same amount. This has the effect of modeling all products with zero observed sales as having the same underlying local demand, whereas the Local RE specification models these products as having different underlying local demand based on the random effects.

5.1 Demand Parameters Constant Across Markets

We begin by discussing the demand parameters that are constant across locations. A summary of our main demand estimates is presented in Tables 5 and 6 for men's and women's shoes, respectively. Within each table, there are four sets of estimates, corresponding to: (1) Local NL-Unadjusted; (2) Local NL-Adjusted; (3) National NL; and (4) Local RE. For each of our specifications, we also compute individual-product-level price elasticities. For national-level estimates, price elasticities are computed as

$$e_j = \frac{\partial \log \pi_j}{\partial \log p_j} = \alpha p_j \left(\frac{1}{1-\lambda} - \frac{\lambda}{1-\lambda} \pi_{j|c} - \pi_j \right),$$

and for local-level estimates, price elasticities are computed as

$$e_j = \frac{\partial \log \sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell j}}{\partial \log p_j} = \frac{\alpha p_j}{\pi_j} \left(\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell j} \left(\frac{1}{1-\lambda} - \frac{\lambda}{1-\lambda} \pi_{\ell j|c} - \pi_{\ell j} \right) \right).$$

Specifications (1) and (2), Local NL-Unadjusted and Local NL-Adjusted, demonstrate the issues that arise given the number of local zeros in the data. The issues are akin to the biases created by truncation and censoring for specifications (1) and (2), respectively.[26] Of particular concern for us are the price coefficients and the nesting parameters. Specification (1) truncates the data at zero, resulting in a selection bias that leads both the nesting parameter and the price coefficient to be biased toward zero. Specification (2), shows the poor performance of the asymptotic adjustment when the fraction of zeros is very large. In particular, the price coefficients become positive for

[26] Standard tools to address truncation/censoring cannot be applied in this setting because the values of the dependent variable are not truly known even for observations that are not truncated/censored. Rather they are implied by the model and, more importantly, depend on the values that are truncated/censored.

both men and women. The 95% of products that are not purchased are all adjusted by the same amount, essentially leading to left censoring of the data, where the mean utility ($\delta_{\ell j}$) is known only to be below this value. The impact of these biases imply price elasticities that are inelastic using unadjusted shares and positive using adjusted shares. Elasticity estimates using unadjusted local shares for mens' and womens' shoes (1) are a quarter of the size of the price elasticities resulting from the National NL (3) and Local RE (4) models. We find a similarly large effect of the selection bias for women.

Overall, the existing approaches using adjusted and unadjusted shares perform poorly. We have also examined additional specifications, such as estimating demand market-by-market. The results are similarly poor because the primary issues stem from the zeros problem, not from the lack of specification flexibility. Note that adding random coefficients would not only be infeasible with millions of observations due to the computational burden, it would also not address the zeros. This is true even after aggregating data. For example, even at the brand level, the choice set is reduced to only a few hundred options, but the number of market-level zeros is still close to 60%.

In the last two columns of Table 5 and Table 6, we report the results from estimating the National NL and Local RE models. The results differ substantially from the preceding two columns. All coefficients are significant and have the correct sign. The nesting parameters for men's shoes are 0.81 and 0.47, suggesting substitution within category is important. For women's shoes, we obtain nesting parameters of 0.37 and 0.28, implying lower substitutability among shoes within category, compared to men. The two models also yield similar (national) price elasticities, (-3.6, -3.3) and (-3.1, -2.1) for men and women, respectively.[27] One thing to note is the parameter estimates are similar. This is not surprising because these are national-level (δ_j) coefficients, and the two models predict the same aggregate demand. However, the local-level predictions of the models are very different because the Local RE model retains information on the distribution of heterogeneity across locations (Table 7) and accounts for product crowding in the spirit of Ackerberg and Rysman (2005). We demonstrate this in the next section.

Turning to the coefficients on our review variables, we can see that the overall rating has the expected sign, with higher ratings having positive effects on demand. The coefficients for look and comfort are also positive, but have smaller effects than the overall rating. Meanwhile, our indicator for no reviews takes on positive signs for both men's and women's shoes. This variable

[27] The empirical literature on shoe demand is limited. Roberts, Yi Xu, Fan, and Zhang (2017) look at imports of Chinese footwear. For the US, their elasticities are slightly smaller; however, their definition of a shoe is broader than our study.

largely captures the demand for new products before there has been an opportunity to review 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.

5.2 Across-Market Heterogeneity

A key advantage of the Local RE model over the National NL model is that it rationalizes the distribution of local demand. As the National NL specification models only aggregate demand, any differences in demand across locations cannot be disentangled. On the other hand, the Local RE specification estimates the same aggregate demand, but explicitly models how demand varies across local markets. Our estimates for the across-market demand heterogeneity in the Local RE model are presented in Table 7.

We find all of the across-market heterogeneity parameters to be statistically significant and precisely estimated, except for two: men's clogs and men's sandals in the winter. More importantly, the parameters that are statistically significant are also highly significant economically. For example, the smallest statistically significant σ_c for men is boots in the summer at 0.21 and for women is clogs at 0.63. To put these numbers in perspective, a one standard deviation increase in a product's draw of $\eta_{\ell j}$ is equivalent to a decline in price of around \$13 and \$52, for men's boots (in summer) and women's clogs, respectively. These parameters represent the variance of across-market demand heterogeneity within a category, suggesting that there is significant variation in demand across markets that goes beyond a simple boots versus sandals story – in fact, there is significant heterogeneity in tastes for all products. These large effects will have important implications for consumer welfare analysis, as we will see in the following section. Finally, although the coefficients appear similar across categories, this is largely due to similar distributions of zero sales across categories, controlling for the differences in the number of products across categories. Additionally, their magnitudes in dollar terms do differ in economically meaningful ways. For example, a one standard deviation increase in a product's draw of $\eta_{\ell j}$ ranges from \$0-34 for men and \$52-73 for women, depending on category.

Across-market demand heterogeneity is important for rationalizing the distribution of local sales in the data. Figure 5 illustrates how σ_c rationalizes the distribution of local sales. For each category, we simulate sales using our Local RE model for two scenarios: (i) assuming our estimated level of across-market demand heterogeneity and (ii) assuming no across-market demand heterogeneity. We see the Local RE model closely follows the observed data, which may not be surprising because the micro moments match local zeros. However, we see that assuming homogeneous demand across markets systematically understates the percentage of local zeros. Given the large number of product-location pairs, these deviations are quite large. For example, under-predicting the percentage of zeros by 0.5 percentage points implies predicting sales for 65,934 men's and 85,622 women's product-location pairs that are observed in the data to be zero.

6 Analysis of the Estimated Model

With our demand estimates, we now conduct a series of counterfactual exercises to quantify the gains from online variety (Section 6.1). We compare consumer surplus and retail revenue under the large (observed) online choice set to the counterfactual surplus and revenue obtained under a limited assortment of products. This mimics a world in which consumers do not have access to online retail. We consider two scenarios: (1) where local assortments target local demand and (2) where local assortments are standardized, which is analogous to the counterfactuals found in the existing literature. Second, by introducing hypothetical price changes, we examine the relative importance to consumers of increased access to variety versus reduced prices (Section 6.2). 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 in our online retail data (Section 6.3).

Implicit in these counterfactuals are a couple of assumptions that are common in the literature. First, we assume the relevant universe of products is observed in our data. Given the size of our retailer's assortment, we believe this is a reasonable assumption. Second, we assume the estimated preferences generalize to brick-and-mortar consumers. If, for example, consumers shop online only if they cannot find their preferred variety locally, niche preferences would be selected into our data. Although we cannot formally test for this, patterns in the data suggest this assumption is reasonable. For example, in Section 2.3, we illustrate that consumers in colder states purchase more boots and fewer sandals. Furthermore, brands that are likely to be available at most brick-and-mortar stores are also popular in the online data.

6.1 The Gains from Increasing Access to Variety

The objective of our main counterfactual is to quantify the increase in consumer surplus and retail revenue from increasing access to variety in the presence of across-market demand heterogeneity. Mechanically, to compute our counterfactuals, we draw a set of ηs for each location. Using these taste draws, along with the recovered national mean utilities, products are then ranked in each location by their location-specific market shares.[28] Products with the highest local shares are included in the counterfactual choice set. These products make up the "pre-internet choice set." For each counterfactual choice set, local-level choice probabilities are then recalculated. Using these probabilities, we simulate location-level purchases, which then allows us to compute counterfactual consumer surplus and retail revenue.

We utilize our local retailer data and information on the number of shoe stores from the US Census County Business Patterns (CBP) to set local assortment cutoffs. Although we cannot directly match our online sales data and our brick-and-mortar assortment data, we can use the counts as a guide for our selection of the local assortment sizes. For each local market, we compute the average number of unique varieties across stores in our Macy's and Payless data. We then multiply that average by the number of local shoe stores observed in the CBP data to get an estimate of the number of unique varieties available to consumers in that location. Because some markets do not contain a Macy's or a Payless, we predict the number of unique varieties based on population so that each market receives the assortment based on the prediced values of

$$\log(a_\ell) = \beta_0 + \beta_1 \log(N_\ell) + \varepsilon_\ell,$$

where a_ℓ is the number of unique varieties from the exercise above and N_ℓ is the size of the local population. For robustness, we also conduct the counterfactuals for a range of thresholds in the following section, which mimic the exercises performed in the previous literature. With the local choice set defined, we define location-level consumer surplus as

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

and retail revenue is defined as

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

where M is the size of the national population (for men and women, respectively).

Table 8 summarizes our main findings and compares estimates across various demand spec-

[28] This assumes that retailers have full information regarding consumer preferences, which may understate the gains from variety as consumers may not have the optimal basket of goods offered locally (Aguiar and Waldfogel 2018).

ifications. Our estimates of the gains from online variety, accounting for across-market demand heterogeneity and tailored pre-internet product assortments are contained in the middle column (Local RE - Tailored Assortment). We estimate consumer surplus gains of $37.1 million, or 5.8%. Our interpretation is that these numbers are sizable, but are about 45% lower than the gains without tailored assortments (Local RE - National Assortment), which is the exercise performed in the existing literature. This implies abstracting from local assortments overstates consumer welfare in our context by 75.4% in absolute terms and 83.5% in percentage terms. The overstatement occurs because the baseline welfare (pre-internet) of consumers is lower when choice sets are determined by national preferences than when they are locally targeted.

Our results have important implications for policy. Previous findings would suggest that antitrust authorities and policy makers should weigh potential variety changes much more heavily than potential price changes. It also implies that there may have been a dramatic decline in the price indices of goods sold online. However, by omitting heterogeneity in preferences across geography and retailer response, the previous literature has overstated the gains from increased product variety by a significant margin. Although we find the gains from online variety to be significant, as we show in the next subsection, the gains from online variety are not significantly greater than the gains from the competitive price effect of e-commerce. This is in sharp contrast to the previous literature which suggests the gains from online variety are enormous relative to the gains from reduced prices.

The Local NL model provides vastly different estimates for the gains from variety compared to the Local RE model, and are found in the first column. Consistent with Ellison and Glaeser (1997), using the unadjusted shares and ignoring the local-level small sample problem exaggerates estimated heterogeneity across markets. By assuming products without an observed sale are completely unwanted at that particular location, the customized counterfactual choice set already satisfies the entirety of local consumer demand and we estimate the consumer welfare gains to be nonexistent. The counterfactuals using adjusted shares are omitted because the positive estimated price coefficient implies nonsensical results.

Finally, by comparing our Local RE model to the National NL model (last column), we can see the effect of failing to account for the variance of the logit error in the spirit of Ackerberg and Rysman (2005). The tendency for logit-style demand models to overstate welfare gains under large changes in the choice set is evident in the National NL results, where estimates of consumer surplus gains are twice that of our estimator with nationally standardized assortments. Thus, an additional benefit of using our Local RE approach is not only does it provide an estimate of the distribution of local demand, but it also reduces the role of the idiosyncratic logit error draws in the analysis.

Our results also have implications concerning assortments at brick-and-mortar retailers. By comparing the results of the nationally standardized assortment with localized assortments, we find revenue is 5.8% higher under the latter. This suggests that there may be a significant incentive for brick-and-mortar stores to target local demand, depending on the potential dis-economies of scale due to localization.

6.2 Consumer Welfare and Retail Prices

In our analysis above, we assume that prices do not change in the absence of the online retail. We do so for two main reasons. First, the existing literature has found the welfare gains from variety are enormous relative to the gains from lower prices. This is despite nontrivial reductions in price due to online retail, estimated to be between 2%-16%, depending on the product category.[29] Second, although it is possible to back out marginal costs from our demand model by specifying a supply-side, the implied costs would be for our online retailer and would not reflect the costs or the structure of competition faced by hypothetical brick-and-mortar retailers. In this subsection, we examine the impact of lower prices and increased variety on consumer welfare using specified price reductions.

Using the results from our preferred specification (Local RE model with tailored assortments), we rerun our counterfactuals for a range of price changes. These price changes are in line with studies that have examined the impact of online competition on prices. More specifically, we take the pre-internet choice sets constructed for our original counterfactuals and now allow for higher pre-internet prices. The entry of the online retailer then gives consumers access to the entire choice set and provides an across-the-board price reduction for all products. We take the reduced prices to be observed prices, so the counterfactual prices are price increases relative to what is observed in the data.

The results of this exercise are presented in Table 9, which compute the total welfare change and the welfare change attributable to the price reduction. We find that the variety effect does not swamp the price effect on consumer welfare even for relatively modest price changes. For

[29]Examples include: 9-16% in books (Brynjolfsson and Smith 2000), 2% in cars (Morton, Zettelmeyer, and Silva-Risso 2001), 8-15% in life insurance (Brown and Goolsbee 2002), 16% in electronics (Baye, Morgan, and Scholten 2003).

example, with a 5% price reduction due to e-commerce, the price reduction accounts for slightly more than half of the total change in consumer welfare. The remaining half comes from increased product variety. Our results differ substantially from the combined conclusions of Brynjolfsson and Smith (2000) and Brynjolfsson, Hu, and Smith (2003), where the ratio of gains from variety versus gains from a 9-16% price reduction are between 7:1 and 10:1 for books. Our results suggest a 10% price reduction yields an analogous ratio of 0.49:1. That is, the price effect actually dominates the variety effect. Although the product categories differ, two factors differentiate our analysis from previous work, which are likely to hold regardless of category: 1) we allow local retailers to tailor their assortments to local demand and 2) we allow the size of the local choice set to differ by the size of the market. The gains from online variety are mitigated because consumers already have access to many of the products that they want. As a consequence, price changes are relatively more important to the overall welfare gain.

6.3 Long Tail Analysis

Our results have important implications for the understanding of the long tail phenomenon observed in online retail. Figure 6 plots the cumulative share of revenue going to the top K products (x-axis) for all observations in the data (national line). The revenue curve features a long tail (fat tail), which the existing literature views as a great source of welfare gain for consumers. The traditional view is that prior to the internet, consumers had access to only a small subset of products, like our counterfactual exercises. However, with the internet, large gains are achieved as consumers switch to new, more preferable varieties made available through e-commerce. One of the key ideas behind the long tail phenomenon is that although the incremental gain from adding an individual variety is small, the sum total of an enormous number of additional varieties is large.

Our article offers a different perspective on the composition and formation of the long tail – that it is the consequence of aggregating over markets with different local demands. To see this, in addition to the national curve, Figure 6 also plots the revenue curve for the following subsets of the data: median market (by number of monthly sales), middle 10% (p45-p55) and middle 50% (p25-p75). For the median market, we can see that there is an extremely short tail, with fewer than 2,000 products making up the entirety of local sales. The next lines (p45-p55, p25-p75) aggregates the sales data for the middle 10% and 50% of markets, respectively. Because the popularity of products varies across geographic markets, aggregating over markets increases the number of different varieties sold and decreases the density of sales among the top ranked products. As markets are combined, sales become less concentrated among the top products producing a lengthening effect of the revenue tail.

The fact that local revenue tails are short suggests that the consumer gains of from additional varieties will be small. In the case where we treat local market shares as being measured without error, we found the gains to be exactly zero. However, one of the key contributions of the methodology presented in this article is to allow measurement error in local shares. With our Local RE model, we can remove the small sample problem at the local level. Figure 7 contains the same median market (data) revenue curve along with the national revenue curve found in Figure 6. We add a line called "Median (Simulated)" which removes the small sample problem for that location. Figure 7 highlights two important results that drive the key findings of the article. First, it suggests local tails are quite a bit longer than suggested by the raw data. This is captured in the welfare analysis by how the Local RE specification models the underlying local demand of products observed to have zero sales. This is in contrast to the case where the zeros are dropped, leading us to conclude the gains from variety are nonexistent, and to the case where all the zeros are adjusted by the same amount, leading to estimates of upward sloping demand curves. Second, the tail is much shorter than the national-level curve, which is consistent with significant across-market demand heterogeneity. By omitting across-market preference heterogeneity, we overstate the value of additional variety by over 75%.

7 Robustness

In this section, we conduct three sets of robustness analysis. We first link across-market demand heterogeneity (σ) with the distribution of local sales and the resulting welfare implications. Next, as our welfare analysis relies on a specified counterfactual choice set, we examine the robustness of our results to the size of the choice set. Finally, we comment on the small samples issue in the data and the long tail phenomenon.

7.1 Across-Market Demand Heterogeneity, Local Zeros, and Welfare

Demand for individual products differs across markets in our model according to our random effects, σ_c. The size of σ_c impacts both the number of local-level zeros and the consumer welfare gains from online variety. Figure 8 shows this relationship. The plot is centered on the estimated σ (x-axis) and corresponding percent of local zero sales and estimated consumer welfare gains (y-axis) of the Local RE model. As across-market heterogeneity increases, corresponding to σ between 100% and 200%, the fraction of zeros increases. The plot demonstrates how the model is identified. There is a monotonic relationship between zeros and across-market demand heterogeneity.

The second feature the plot highlights is the relationship between across-market demand heterogeneity and the gains from online retail. As σ_c increases, each location has stronger preferences for a smaller subset of products. Because this makes it easier for local retailers to cater to these preferences, the additional value created by the large online choice set is smaller. Although we use shoes as our example good, the internet has increased product variety broadly. Figure 8 illustrates the central role across-market demand heterogeneity plays in the calculation of the gains from online variety. In the extreme where there is no across-market demand heterogeneity, the gains are nearly 60% higher relative to our baseline estimates for shoes (i.e. on the right y-axis: 160 compared to 100). If heterogeneity were twice that of our estimates, the gains from online variety would be about 35% the size of our baseline estimates. The gains from additional variety fall rapidly as the degree of across-market demand heterogeneity rises. Ultimately, Figure 8 highlights the importance in this type of analysis of accounting for the differences in local demand and the targeting of local assortments.

7.2 Welfare and Counterfactual Choice Sets

Like the previous studies, our results are based on a specified counterfactual choice set. Here we conduct sensitivity analysis to the choice set size. Our general finding is that the absolute size of the overstatement is sensitive to the size of the counterfactual assortment size, but the percentage overstatement is fairly robust across a wide range of threshold sizes and in line with our findings from the previous section.

Table 10 presents the change in consumer welfare and the size of the overstatement resulting from various thresholds of the counterfactual choice set. The counterfactual choice sets are specified to be: mean baseline choice set, 3,000, 6,000, 12,000, and 24,000 total varieties, split between men's and women's shoes. For comparison, we also include our baseline results from the previous section in the top row. Unsurprisingly, as the size of the counterfactual choice set increases, the gain consumers derive from access to the remaining products decreases. This decrease occurs faster under locally-customized assortments than under a nationally standardized assortment. As a result, the percentage overstatement tends to increase in the assortment size, despite the absolute size of the overstatement decreasing. This pattern is illustrated in Figure 9, which 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 (dashed). The absolute overstatement peaks at $100 million with about about 5% of products, whereas the percentage overstatement approaches infinity as the fraction of available products approaches one.

Table 11 presents the change in retail revenue at various thresholds of the counterfactual choice set. With retail revenue we find that as assortment sizes increase, the gain from customizing assortments to local demand decreases in size. However, a typical large brick-and-mortar shoe retailer stocks, at most, a few thousand varieties. Our results imply that a national retailer stocking 3,000 products in each store could increase its revenue by about 28% by moving to a locally customized inventory from a nationally standardized one. This suggests that there may be significant incentives for large national brick-and-mortar shoe retailers to customize their assortments to local demand.

Figure 10 graphs the increase in retail revenue due to local customization of assortments, measured in millions of dollars (solid) and as a percentage (dashed). The percentage increase monotonically decreases 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. The absolute gain in revenue from localization peaks at $120 million at about 10% of products.

7.3 Small Sample Sizes and the Long Tail

We may be concerned that the long tail observed in our aggregated data is actually due to small sample sizes at the local level, rather than driven by across-market demand heterogeneity. Figure 11 graphs the cumulative share of revenue going to the top K products for the median CSA, middle 10% (p45-p55), middle 50% (p25-p75), and the national-level markets across four panels (solid). To test how sampling impacts the revenue curve, we remove all products in which only a single local sale occurs (dashed).

As expected, we find that removing single sale products shortens the revenue tail. For the median market (a), the already extremely short tail shortens further. For the middle 10% of markets (b) the shortening is quite large, but this effect diminishes substantially with aggregation to the middle 50% (c). In particular, at the national level (d) we still obtain a long tail pattern, even with all of the single sale products removed at the local level. This suggests that aggregation does, in fact, average out the effects of small sample sizes and gives us confidence that our long tail results are not driven by one-off purchases.

8 Conclusion

In this article, we quantify the effect of increased access to variety due to online retail on consumer welfare and firm revenue. To perform this analysis, we develop new methodology that allows us to investigate across-market demand heterogeneity when market shares are measured with error. The methodology relies on a large number of markets and products, both of which are common in big data sets. Applying the method to novel data on online shoe sales, we find products face substantial heterogeneity in demand across markets, and that this heterogeneity helps explain the distribution of sales we see in the data.

The presence of across-market demand heterogeneity has important implications for both firm strategy and consumer welfare. On the supply side, differences in local demand may create an incentive for retailers to tailor assortments and our brick-and-mortar data suggests that local shoe stores are reacting to these incentives. Our results suggest local retailers may generate about 5.8% additional revenue by targeting local consumers. On the demand side, there are several potential avenues through which online retail benefits consumers. Although the early literature focused on the competitive pressure online retail placed on prices, the variety channel has since been singled out as a much larger source of consumer welfare gains in the context of online retailing. However, by estimating across-market demand heterogeneity, accounting for sampling and controlling for the tendency of logit-style models to inflate the value of additional products, our analysis suggests that the variety channel is substantially less important than previously thought. Our estimates suggest the value of increased product variety due to e-commerce is about 45% lower than the existing studies.

Although we bring in new, rich data and propose new methodology to estimate demand with a significant number of local zeros, our results are subject to some limitations. First, like the existing literature, we do not see the arrival of e-commerce and the associated increase in product availability in the sample. Thus, we cannot trace out the price and variety effects of online retail over time. Instead, our results come from a retrospective analysis, where we first estimate demand and then construct a hypothetical world, without e-commerce, by reducing the choice set available to consumers. Because we observe only one retailer, we do not formally model assortment and price competition of local retailers. However, we conduct a number of counterfactual scenarios that suggest the variety effects due to e-commerce are not vastly greater than the plausible price effects. Our data also requires us to abstract from consumer search, which can be an important feature of both offline and online retail. If online search is more costly than offline search, our results provide an upper bound on the welfare gains, however, if the reverse is true, we understate the gains by the additional cost of searching for varieties at local retailers.

Finally, our analysis focuses on a commonly purchased item, shoes. Although there is strong evidence that geographic preference heterogeneity exists across a wide swath of categories (Bronnenberg, Dhar, and Dube 2009), its importance may differ across categories. It would be interesting to examine the degree of across-market demand heterogeneity and the value of variety over a broad set of categories. The value of increased product variety for these categories undoubtedly depends on local retailer responses to this heterogeneity.

References

Ackerberg, D. A., and M. Rysman (2005): "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects," RAND Journal of Economics, 36(4), 771–788.

Aguiar, L., and J. Waldfogel (2018): "Quality Predictability and the Welfare Benefits from New Products: Evidence from the Digitization of Recorded Music," Journal of Political Economy, 126(2), 492–524.

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

Arkolakis, C., S. Demidova, P. J. Klenow, and A. Rodriguez-Clare (2008): "Endogenous Variety and the Gains from Trade," American Economic Review, 98(2), 444–50.

Baye, M. R., J. Morgan, and P. Scholten (2003): "The Value of Information in an Online Consumer Electronics Market," Journal of Public Policy & Marketing, 22(1), 17–25.

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), 841–890.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," Journal of Political Economy, 112(1), 68–105.

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.

BRODA, C., AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," The Quarterly Journal of Economics, 121(2), 541–585.

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.

BROWN, J. R., AND A. GOOLSBEE (2002): "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," Journal of Political Economy, 110(3), 481–507.

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.

BRYNJOLFSSON, E., AND M. D. SMITH (2000): "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," Management science, 46(4), 563–585.

CARDELL, N. S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," Econometric Theory, 13(2), 185–213.

CHINTAGUNTA, P. K., AND J.-P. DUBE (2005): "Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data," Journal of Marketing Research, 42(3), 368–379.

CHOI, J., AND D. R. BELL (2011): "Preference Minorities and the Internet," Journal of Marketing Research, 48(4), 670–682.

CHOW, Y., AND T. LAI (1973): "Limiting Behavior of Weighted Sums of Independent Random Variables," The Annals of Probability, pp. 810–824.

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, 67(3), 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., Z. LU, AND X. SHI (2013): "Estimating Demand for Differentiated Products with Error in Market Shares," Working paper.

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

GOOLSBEE, A., AND A. PETRIN (2004): "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," Econometrica, 72(2), 351–381.

HWANG, M., B. J. BRONNENBERG, AND R. THOMADSEN (2010): "An Empirical Analysis of Assortment Similarities Across US Supermarkets," Marketing Science, 29(5), 858–879.

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(2), 132–157.

Morton, F. S., F. Zettelmeyer, and J. Silva-Risso (2001): "Internet Car Retailing," The Journal of Industrial Economics, 49(4), 501–519.

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

Roberts, M. J., D. Yi Xu, X. Fan, and S. Zhang (2017): "The Role of Firm Factors in Demand, Cost, and Export Market Selection for Chinese Footwear Producers," The Review of Economic Studies, forthcoming.

Romer, P. (1994): "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions," Journal of Development Economics, 43(1), 5–38.

Thomas, C. (2011): "Too Many Products: Decentralized Decision Making in Multinational Firms," American Economic Journal: Microeconomics, 3(1), 280–306.

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 Proofs

Proposition 1. *For each product $j \in J$, applying the law of large numbers (for weighted sums with population weights ω_ℓ^L) in L and integrating out over η gives*

$$\sum_{\ell=1}^{L} \omega_\ell^L \pi_{\ell j}(\eta_\ell; \delta, \lambda) - \pi_j \to_{a.s.} 0 \tag{3}$$

Proof. By construction of the local-choice probability, $\pi_{\ell j}(\eta_\ell; \delta, \lambda) \in (0, 1)$ for all $j \in J \cup \{0\}$ and for all $\eta_\ell \in \mathbb{R}^J$. Thus, each random variable is bounded below by zero and above by one, and hence, both $E[\pi_{\ell j}(\eta_\ell; \delta, \lambda)] < 1 < \infty$ and $Var[\pi_{\ell j}(\eta_\ell; \delta, \lambda)] < 1 < \infty$. Therefore,

$$\sum_{\ell=1}^{L} \frac{Var[\pi_{\ell j}(\eta_\ell; \delta, \lambda)]}{\ell^2} < \sum_{\ell \in L} \frac{1}{\ell^2} < \infty.$$

We define the weights for a given L to be $\omega_\ell^L = \frac{w_\ell}{\sum_{\ell=1}^L w_\ell}$ so that $\sum_{\ell=1}^L \omega_\ell^L = 1$. To apply the law of large numbers for weighted sums of Chow and Lai (1973), we need there to exist a sequence $(w_\ell)_{\ell=1}^\infty$ such that $\sum_{\ell=1}^\infty w_\ell^2 < \infty$. The defined weights clearly satisfy this because $\sum_{\ell=1}^\infty (w_\ell^\infty)^2 < \left(\sum_{\ell=1}^\infty w_\ell^\infty\right)^2 = 1$. Therefore,

$$\sum_{\ell=1}^L \omega_\ell^L \pi_{\ell j}(\eta_\ell; \delta, \lambda) - \sum_{\ell=1}^L \omega_\ell^L E\left[\pi_{\ell j}(\eta_\ell; \delta, \lambda)\right] \to_{a.s.} 0.$$

Note that $\pi_{\ell j}(\eta_\ell; \delta, \lambda)$ differ across locations only by their draw of η_ℓ and that each location is identically distributed. Thus, the expected value of $\pi_j(\cdot)$ is equal across locations, and equal to the aggregated choice probability. That is,

$$\sum_{\ell=1}^L \omega_\ell^L E\left[\pi_{\ell j}(\eta_\ell; \delta, \lambda)\right] = E\left[\pi_{\ell j}(\eta_\ell; \delta, \lambda)\right] = \pi_j,$$

and our result follows. ■

Derivation of Equation 4: The market share inversion for our random effects model with L locations is

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

Proof. We will find it convenient to write shares as a fraction of the category share. By Bayes' rule, location-specific market shares are

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

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 and the unobserved taste preference across locations. This means the probability of choosing the outside good at location ℓ is equal to

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

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, $D_{\ell c} = \left(\frac{\pi_{\ell c}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda}}$, and plugging this into the aggregate choice probabilities gives

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

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

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

. ■

Proposition 2. *Suppose the law of large numbers applies, i.e.* $\frac{1}{J_c}\sum_{j \in c} \exp\left\{(\delta_j + \eta_{\ell j})/(1-\lambda)\right\}$ *converges in distribution to a constant for each c, then*

$$E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\} \middle| \pi_{\ell c}, \pi_{\ell 0}\right] \to_d E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right], \quad as \quad J \to \infty.$$

Proof. The proof has two components. The first is to take a monotonic transformation of the conditional expectation. With this transformation, we establish that the expectation converges to a constant. We then show the constant must be the unconditional expectation.

Recall that $D_c = \sum_{j \in c} \exp\{(\delta_j + \eta_{\ell j})/(1-\lambda)\}$.

$$\pi_{\ell 0} = \frac{1}{1 + \sum_{c'} D_{\ell c}^{1-\lambda}} \quad \text{and} \quad \pi_{\ell c} = \frac{D_{\ell c}^{1-\lambda}}{1 + \sum_{c'} D_{\ell c}^{1-\lambda}}.$$

To start, we rewrite the condition variable $\pi_{\ell c}$.

$$\pi_{\ell c} = \frac{\left(\frac{1}{J}\frac{J_c}{J_c} D_{\ell c}\right)^{1-\lambda}}{\left(\frac{1}{J}\right)^{1-\lambda} + \left(\frac{1}{J}\frac{J_1}{J_1} D_{\ell 1}\right)^{1-\lambda} + \ldots + \left(\frac{1}{J}\frac{J_c}{J_c} D_{\ell c}\right)^{1-\lambda}} = \frac{\left(\frac{J_c}{J}\frac{1}{J_c} D_{\ell c}\right)^{1-\lambda}}{\left(\frac{1}{J}\right)^{1-\lambda} + \left(\frac{J_1}{J}\frac{1}{J_1} D_{\ell 1}\right)^{1-\lambda} + \ldots + \left(\frac{J_c}{J}\frac{1}{J_c} D_{\ell c}\right)^{1-\lambda}}.$$

Next, we apply a monotonic transformation of the category share:

$$\bar{\pi}_{\ell c} = \left(\frac{\left(\frac{J_c}{JJ_c} D_{\ell c}\right)^{1-\lambda}}{\left(\frac{1}{J}\right)^{1-\lambda} + \left(\frac{1}{J}\frac{J_1}{J_1} D_{\ell 1}\right)^{1-\lambda} + \ldots + \left(\frac{1}{J}\frac{J_c}{J_c} D_{\ell c}\right)^{1-\lambda}}\right)^{\frac{1}{1-\lambda}} = \frac{\frac{J_c}{J}\frac{1}{J_c} D_{\ell c}}{\left(\left(\frac{1}{J}\right)^{1-\lambda} + \left(\frac{J_1}{J}\frac{1}{J_1} D_{\ell 1}\right)^{1-\lambda} + \ldots + \left(\frac{J_c}{J}\frac{1}{J_c} D_{\ell c}\right)^{1-\lambda}\right)^{\frac{1}{1-\lambda}}}.$$

Let $\Psi_\ell(J;c)$ denote the denominator in the sum above, i.e.

$$\bar{\pi}_{\ell c} = \frac{\frac{J_c}{J}\frac{1}{J_c} D_{\ell c}}{\Psi_\ell(J;c)}.$$

Next, rewriting the expectation, we obtain

$$E\left[\frac{\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}}{\Psi_\ell(J;c)} \cdot \Psi_\ell(J;c) \middle| \bar{\pi}_{\ell c}, \pi_{\ell 0}\right] = E\left[\frac{1}{J_c}\sum_{j \in c} \frac{\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}}{\Psi_\ell(J;c)} \cdot \Psi_\ell(J;c) \middle| \bar{\pi}_{\ell c}, \pi_{\ell 0}\right].$$

We can rewrite this conditional expectation using the Law of Iterated Expectations and using that η is i.i.d. within c to obtain

$$E\left[\frac{J_c}{J_c}\frac{1}{J}\sum_{j \in c} \frac{\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}}{\Psi_\ell(J;c)} \cdot \Psi_\ell(J;c) \middle| \bar{\pi}_{\ell c}, \pi_{\ell 0}\right] = E\left[\bar{\pi}_{\ell c} \cdot E\left[\Psi_\ell(J;c) \middle| \bar{\pi}_{\ell c}\right] \middle| \pi_{\ell 0}\right].$$

We start with the inner expectation and show it converges to a number.

We start with $\Psi_\ell(J;c)$. Here we show the conditional variance converges to zero as $J \to \infty$. We state this as a lemma.

Lemma:

$$Var\left(\Psi_\ell(J;c) \,|\, \bar{\pi}_{\ell c}\right) \to_p 0 \quad \text{as } J \to \infty.$$

Proof. By the definition of variance,

$$E\left[Var\Big(\Psi_\ell(J;c)\,|\,\bar{\pi}_{\ell c}\Big)\right] = E\left[E\left[\Psi_\ell(J;c)^2\,|\,\bar{\pi}_{\ell c}\right]\right] - E\left[\Big(E[\Psi_\ell(J;c)\,|\,\bar{\pi}_{\ell c}]\Big)^2\right].$$

By Jensen's Inequality,

$$\begin{aligned} E\left[E\left[\Psi_\ell(J;c)^2\,|\,\bar{\pi}_{\ell c}\right]\right] - E\left[\Big(E[\Psi_\ell(J;c)\,|\,\bar{\pi}_{\ell c}]\Big)^2\right] &\leq E\left[E\left[\Psi_\ell(J;c)^2\,|\,\bar{\pi}_{\ell c}\right]\right] - E\left[E[\Psi_\ell(J;c)\,|\,\bar{\pi}_{\ell c}]\right]^2 \\ &= Var(\Psi_\ell(J;c)), \end{aligned}$$

by the Law of Iterated Expectations. Note that $Var(\Psi_\ell(J;c)) \to_p 0$ by applying the law of large numbers to the weighted averages inside $\Psi_\ell(\cdot)$, i.e. $\frac{1}{J_c} D_{\ell c} \to_d \delta^c$ for each c.[30] ■

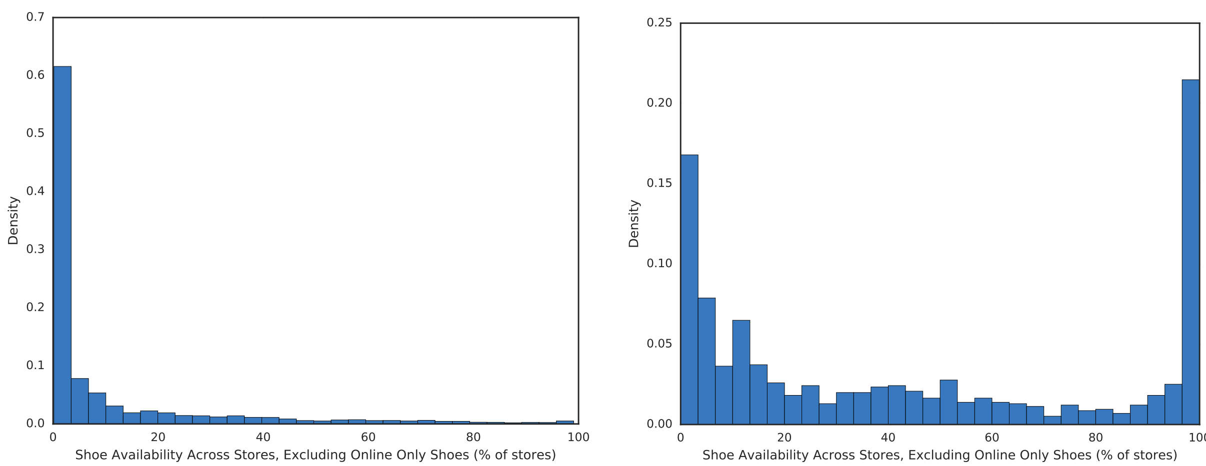
Thus $E\left[\Psi_\ell(J;c)\middle|\bar{\pi}_{\ell c}\right]$ converges to a constant and $\bar{\pi}_{\ell c} \cdot E\left[\Psi_\ell(J;c)\middle|\bar{\pi}_{\ell c}\right]$ converges to a constant by the law of large numbers. By an analogous argument to the lemma, $E\left[\bar{\pi}_{\ell c} \cdot E\left[\Psi_\ell(J;c)\middle|\bar{\pi}_{\ell c}\right] \,\middle|\, \pi_{\ell 0}\right]$ must converge as well. Finally, we have to show it converges to the unconditional expectation of $\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}$. This is immediate by the definition of the expectation. Thus,

$$E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\} \,\middle|\, \pi_{\ell c}, \pi_{\ell 0}\right] \to_d E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] \quad \text{as } J \to \infty.$$

■

B Figures and Tables

Figure 1: Shoe Availability Across Stores



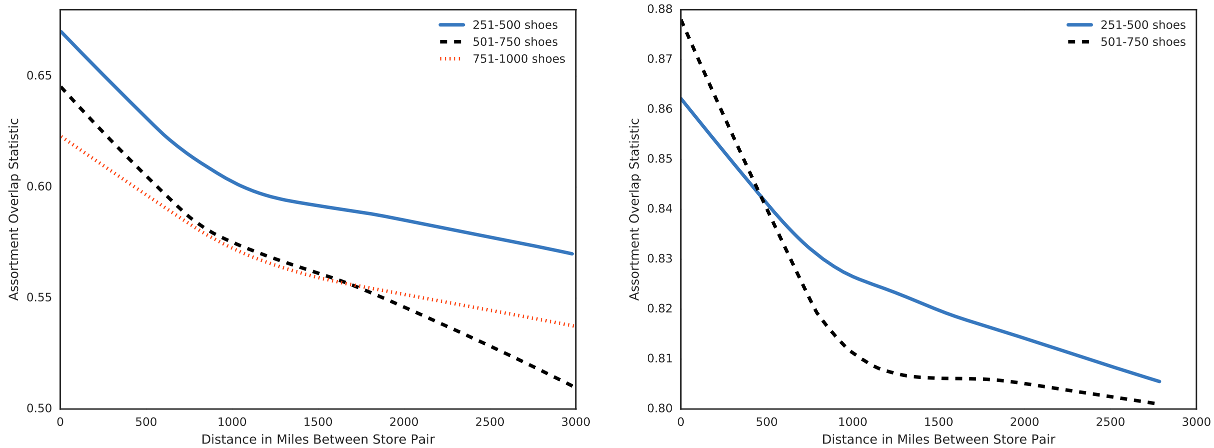
(a) Macy's

(b) Payless

Notes: Histograms showing the density of shoe availability across stores for Macy's and Payless. These histograms exclude online only shoes. The number 100 means that the shoe is available at all stores in the network.

[30] Several assumptions can give this result. For example, we could apply Kolmogorov's two-series theorem under restrictions of the means and variances of independent random variables. Alternatively, we could specify $\{\delta_j\}$ coming from a finite set each of which occurs infinitely often.

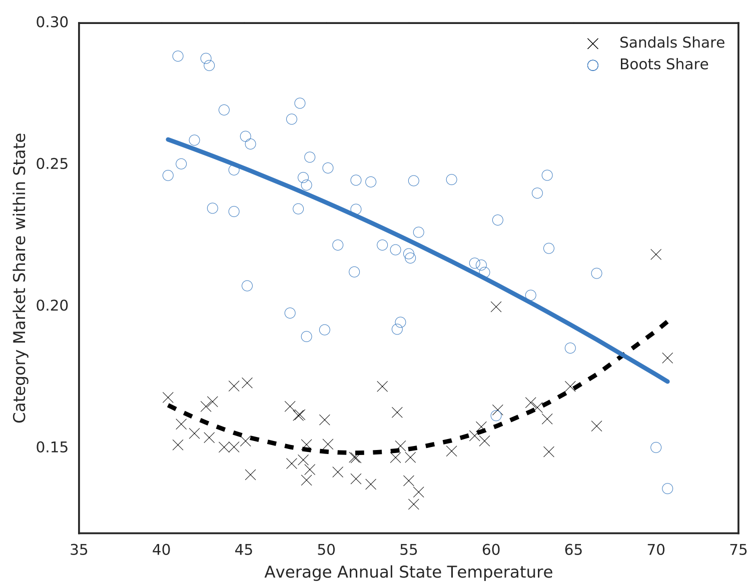
Figure 2: Assortment Overlap by Distance



(a) Macy's (b) Payless

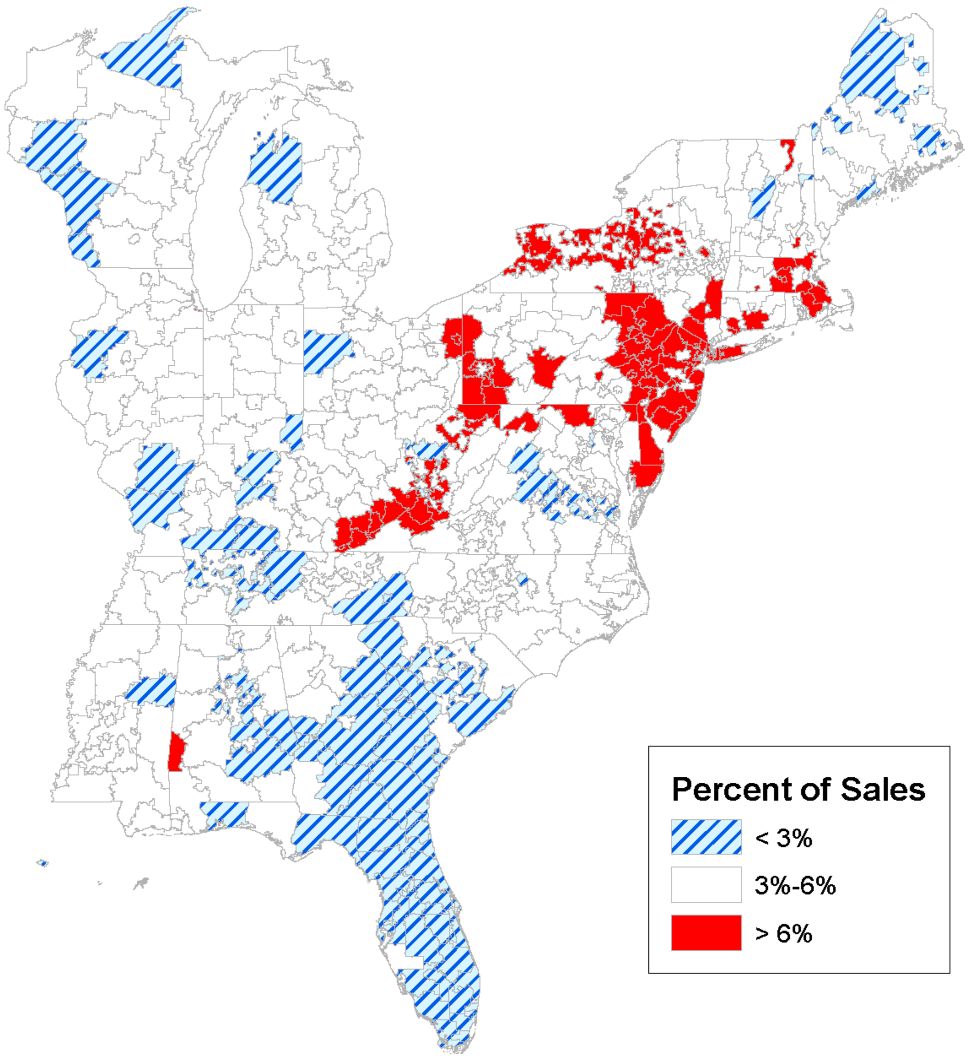
Notes: Lowess fitted values of assortment overlap across stores in the network. Analysis split across stores with similar assortment sizes.

Figure 3: Boots vs. Sandals: Revenue by Temperature



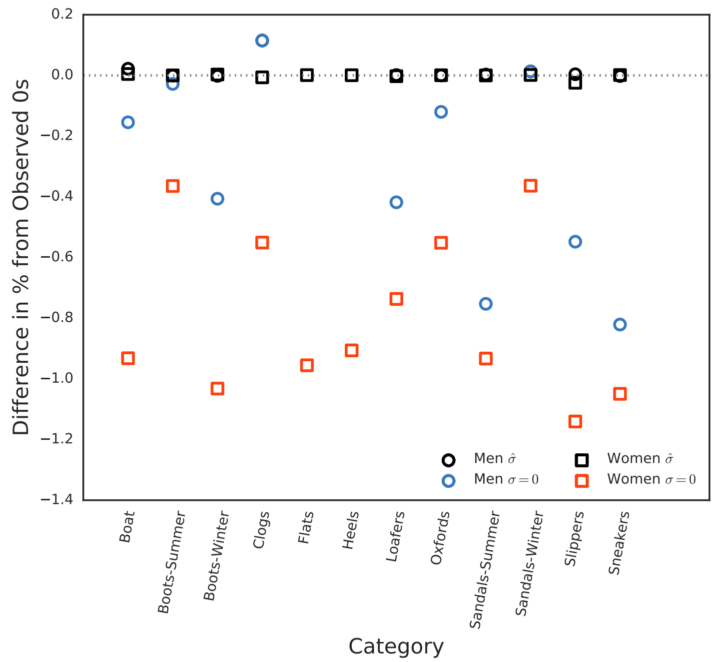
Notes: Scatter plot of average annual state temperatures on the sales of boots and sandals as a share of state revenue. Lines are fitted values of a polynomial regression of degree two.

Figure 4: Revenue Share of a Popular Brand Across Zip3s



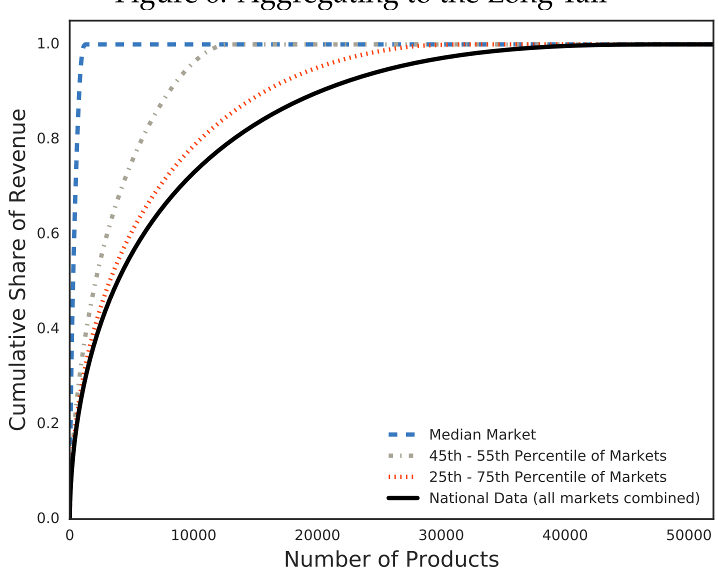
Notes: Map of Eastern US Zip3s – the first 3 digits of a 5-level zip code. The color of the Zip3 corresponds to the local revenue share of a popular brand in the data set. Sales of the brand are concentrated in the Eastern US.

Figure 5: Predicted Zeros Without Across-Market Demand Heterogeneity



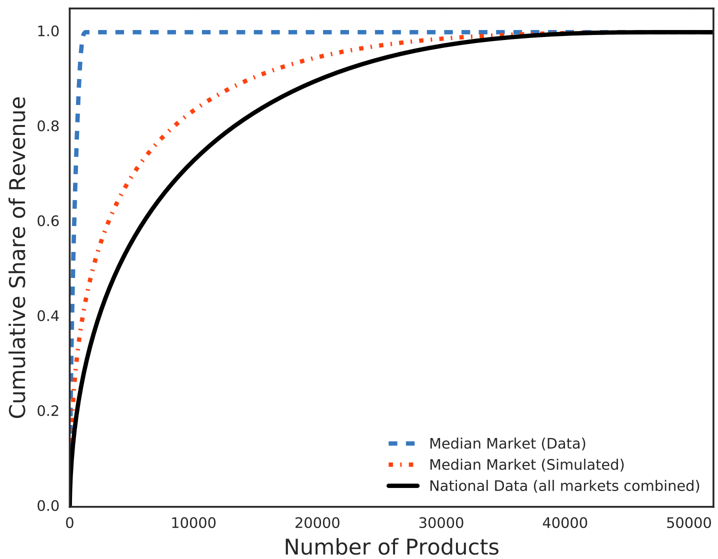
Notes: For each product category, the difference between the observed percentage of zeros and the predicted percentage of zeros. Predicted zeros come from simulation of sales using the estimated level of across-market demand heterogeneity, $\hat{\sigma}$, and assuming no across-market demand heterogeneity, $\sigma = 0$.

Figure 6: Aggregating to the Long Tail



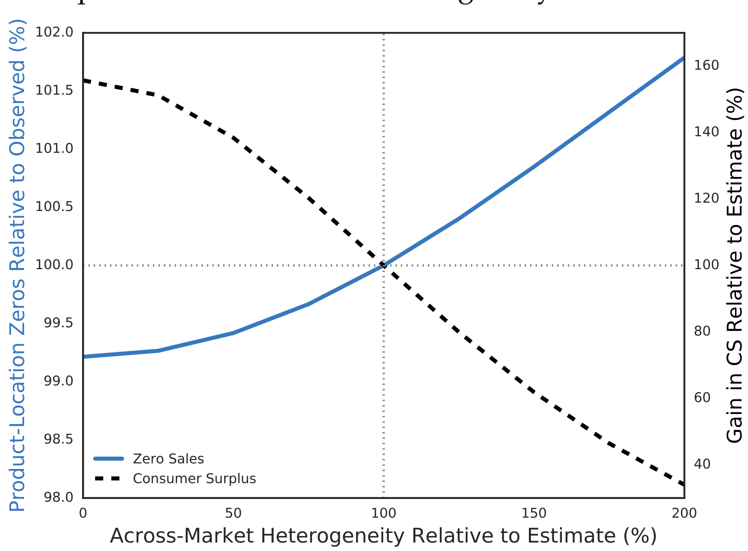
Note: The cumulative revenue curve for the online retailer for different levels of data aggregation. The index of products corresponds to sales rank.

Figure 7: Local Tail: Correcting for Small Samples



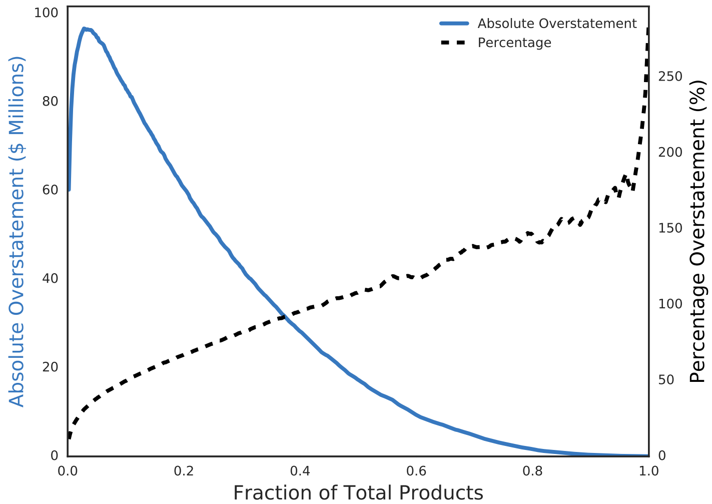
Notes: For the median local market, the cumulative share of revenue going to the top products, as seen in the data (dot) and simulated using our estimated demand system (dash-dot). For comparison, we also include the national-level revenue distribution (solid).

Figure 8: Impact of Across-Market Heterogeneity on Zeros and Welfare



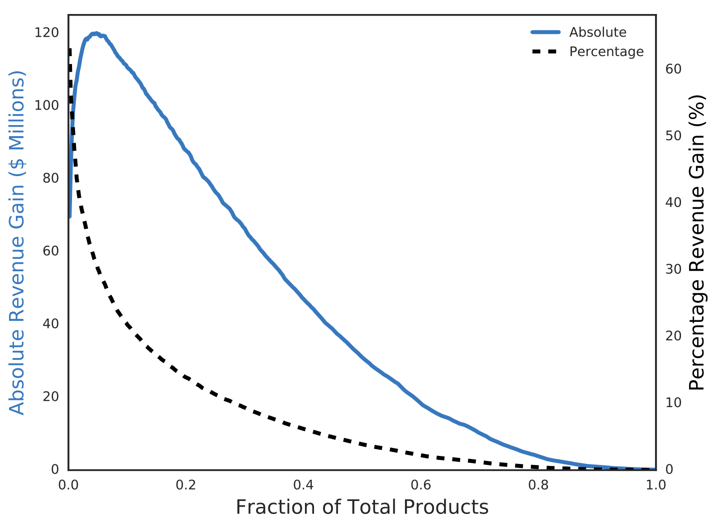
Notes: The predicted number of product-location zeros and estimated consumer welfare for different levels of σ_c. Along the x-axis, "0" indicates no across-market demand heterogeneity, "100" corresponds to our estimates, and "200" corresponds with two times our estimated σ_c.

Figure 9: Overestimation of Consumer Welfare Gains



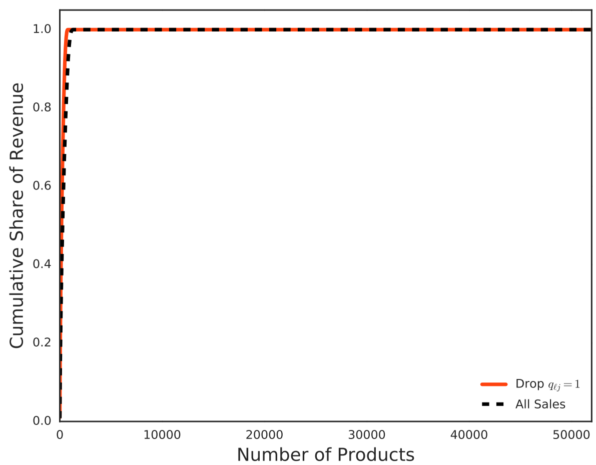
Notes: The overstatement in consumer surplus gains, by counterfactual assortment size, when assuming a nationally standardized assortment vs. a locally customized assortment measured in dollars (red, dotted) and percentage (black, solid).

Figure 10: Increase in Retail Revenue from Localized Assortments

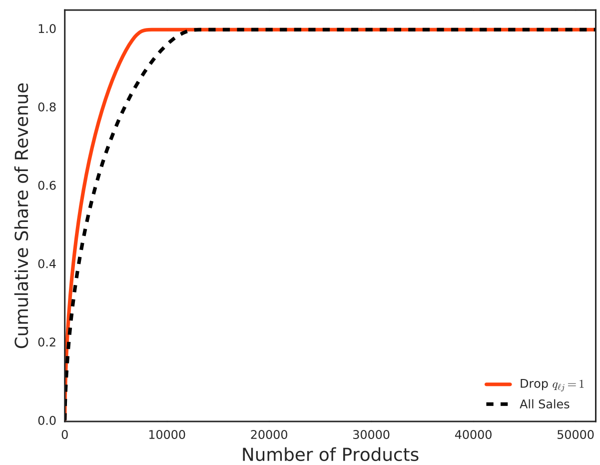


Notes: The gain in retail revenue, by local retailer assortment size, when moving from a nationally standardized assortment to a locally customized assortment measured in dollars (red, dotted) and percentage (black, solid).

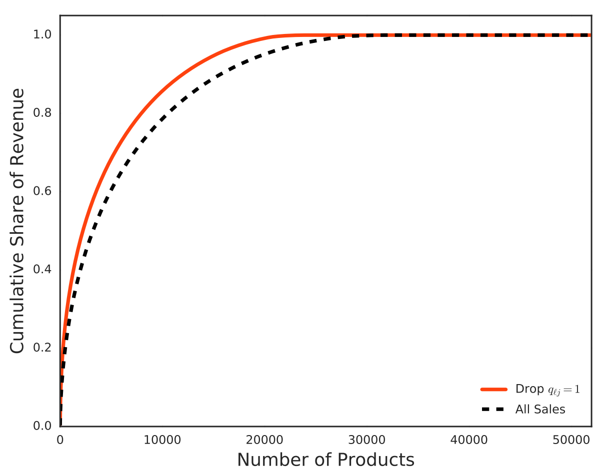
Figure 11: Demand Aggregation Dropping Single Sale Observations.



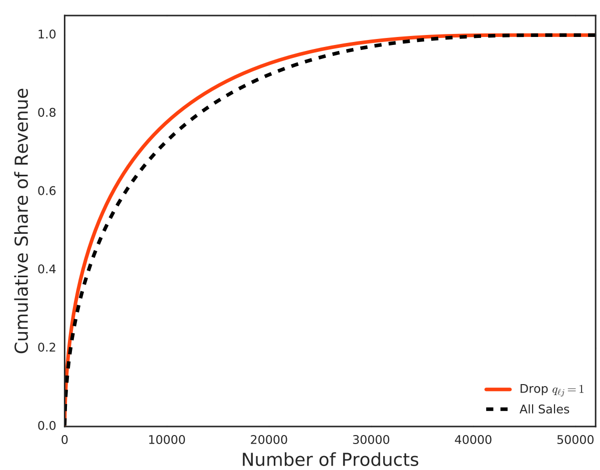
(a) Median Market



(b) 45^{th} - 55^{th} Percentile Markets



(c) 25^{th} - 75^{th} Percentile Markets



(d) Aggregation of All Markets (National)

Notes: For varying levels of aggregation, the cumulative share of revenue going to the top products as seen in the data (dashed line) and after dropping all local-market-level single sales (solid line).

Table 1: Summary of Brick-and-Mortar Data

	Macy's	Payless Shoes
Number of stores	649	3,141
Number of products	13,914	1,430
Percent online exclusive	42.1%	19.2%
Avg. assortment size	871.7	513.0
	(407.9)	(58.4)

Notes: Data collected through macys.com and payless.com. For every shoe-size combination, we check to see if the product is in stock. We then aggregate over sizes to determine assortments. Before aggregation, the total number of observations is $N_{\text{Macy's}} = 93,602,700$ and $N_{\text{Payless}} = 69,451,866$.

Table 2: Local-National Revenue Share Comparison and Multinomial Tests

Market Definition	Number of Markets	Market Top 500 Market	Market Top 500 National	Multinomial Tests - Reject. Rates (%)
5-Digit Zip Code	35,279	99.96	4.69	40.14
3-Digit Zip Code	894	85.12	6.28	65.02
CSA + State	213	67.05	7.23	78.30
CSA	165	70.31	7.19	88.05
State (plus DC)	51	30.04	9.86	94.26
Census Region	4	16.36	14.76	92.86
National	1	15.54	15.54	–

Notes: Multinomial tests: Define $s = \{s_1, ..., s_J\}$ and $s_\ell = \{s_{\ell 1}, ..., s_{\ell J}\}$, then the null hypothesis is $H_0 : s = s_\ell$. CSA + State includes the 165 CSAs and 48 States. NJ and RI are contained in CSAs.

Table 3: Data Disaggregation: The Zeros Problem

Market Definition	Number of Markets	Percent of Zero Sales: Week	Month	Quarter	Annual
5-Digit Zip Code	35,279	99.99	99.96	99.91	99.78
3-Digit Zip Code	894	99.57	98.57	97.07	94.09
CSA + State	213	98.43	95.54	91.98	86.12
CSA	165	98.50	95.80	92.53	87.15
State (plus DC)	51	94.23	85.25	76.27	64.26
Census Region	4	59.83	33.70	21.72	12.17
National	1	28.30	9.27	4.50	1.01

Notes: Percent of products observed to have zero sales, where a product is a SKU.

Table 4: Revenue Share of Top Products with Product Aggregation

Product Definition	Percent Zero Sales	Rev. Share - Top 500 Market	Rev. Share - Top 500 National	Rev. Share - Top 10 Market	Rev. Share - Top 10 National
SKU (shoe + style)	95.54	67.05	7.23	7.59	0.50
Shoe	93.10	73.39	19.07	9.10	2.18
Brand	59.27	–	–	33.91	25.48

Notes: Time horizon fixed at the monthly level and geographies aggregated to the CSA-State level.

Table 5: Demand Estimates with Adjusted Shares - Men's

	Local NL Unadjusted (1)	Local NL Adjusted (2)	National NL Adjusted (3)	Local RE Adjusted (4)
Price	-0.007^{***}	0.001^{***}	-0.006^{***}	-0.016^{***}
	(0.0003)	(0.000004)	(0.0003)	(0.001)
Comfort	0.033^{***}	-0.003^{***}	0.002	0.006
	(0.003)	(0.00006)	(0.003)	(0.008)
Look	0.0004	-0.003^{***}	0.007	0.019^{*}
	(0.003)	(0.00008)	(0.004)	(0.010)
Overall	0.045^{***}	0.0001^{*}	0.059^{***}	0.162^{***}
	(0.003)	(0.00007)	(0.004)	(0.008)
No Review	0.266^{***}	-0.041^{***}	0.274^{***}	0.712^{***}
	(0.012)	(0.0004)	(0.022)	(0.049)
λ	0.103^{***}	0.989^{***}	0.814^{***}	0.465^{***}
	(0.010)	(0.00002)	(0.006)	(0.005)
σ	—	—	—	*
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
Price Elast.				
Mean	−0.876	–	−3.635	−3.333
Std. Dev.	0.619	–	2.735	2.508

Notes: Estimated at the monthly level. "Local NL" (1) estimates nested logit demand at the CSA-State level with unadjusted shares, and (2) estimates nested logit demand at the CSA-State level with Gandhi, Lu, and Shi (2014) adjusted shares. These create local-product-level fixed effects. "National NL" (3) estimates nested logit demand at the national level with Gandhi, Lu, and Shi (2014) adjusted shares, creating national-product-level fixed effects. Finally, "Local RE" (4) 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. Robust standard errors in parentheses.
* estimates for across-market heterogeneity in specification (4) are in Table 7

Table 6: Demand Estimates with Adjusted Shares - Women's

	Local NL Unadjusted (1)	Local NL Adjusted (2)	National NL Adjusted (3)	Local RE Adjusted (4)
Price	-0.005^{***}	0.003^{***}	-0.015^{***}	-0.012^{***}
	(0.0001)	(0.000007)	(0.0004)	(0.0004)
Comfort	0.027^{***}	-0.008^{***}	0.045^{***}	0.050^{***}
	(0.002)	(0.0001)	(0.007)	(0.007)
Look	0.006^{***}	0.003^{***}	0.025^{***}	0.019^{***}
	(0.002)	(0.0001)	(0.008)	(0.007)
Overall	0.047^{***}	-0.012^{***}	0.134^{***}	0.148^{***}
	(0.002)	(0.0001)	(0.009)	(0.008)
No Review	0.263^{***}	-0.158^{***}	0.781^{***}	0.782^{***}
	(0.009)	(0.001)	(0.032)	(0.030)
λ	-0.034^{***}	0.932^{***}	0.370^{***}	0.281^{***}
	(0.003)	(0.00004)	(0.010)	(0.010)
σ	—	—	—	$*$
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
Price Elast.				
Mean	−0.592	–	−3.055	−2.129
Std. Dev.	0.474	–	2.672	1.862

Notes: Estimated at the monthly level. "Local NL" (1) estimates nested logit demand at the CSA-State level with unadjusted shares, and (2) estimates nested logit demand at the CSA-State level with Gandhi, Lu, and Shi (2014) adjusted shares. These create local-product-level fixed effects. "National NL" (3) estimates nested logit demand at the national level with Gandhi, Lu, and Shi (2014) adjusted shares, creating national-product-level fixed effects. Finally, "Local RE" (4) 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. Robust standard errors in parentheses.
$*$ estimates for across-market heterogeneity in specification (4) are in Table 7

Table 7: Parameter Estimates of Across-Market Heterogeneity: $\sigma_j = h(\cdot)$

	Men (1)	Women (2)
Boat	0.257***	0.650***
	(0.046)	(0.022)
Boots - Summer	0.210***	0.836***
	(0.078)	(0.021)
Boots - Winter	0.511***	0.815***
	(0.012)	(0.017)
Clogs	0.000001	0.625***
	(15.783)	(0.016)
Flats	—	0.857***
		(0.017)
Heels	—	0.878***
		(0.017)
Loafers	0.489***	0.772***
	(0.013)	(0.017)
Oxfords	0.282***	0.815***
	(0.042)	(0.024)
Sandals - Summer	0.500***	0.781***
	(0.010)	(0.014)
Sandals - Winter	0.003	0.777***
	(95.029)	(0.017)
Slippers	0.373***	0.808***
	(0.023)	(0.019)
Sneakers	0.544***	0.746***
	(0.008)	(0.013)

Notes: Parameter estimates correspond to "∗", column 3, in Table 5 and Table 6, respectively. Parameters estimated jointly, by gender, with robust standard errors in parentheses. There are no products classified as men's flats or men's heels in the data sample.

Table 8: Welfare Gains From Increasing Variety

	Local NL		Local RE				National NL	
	Unadjusted Shares		Tailored Assort.		National Assort.		National Assort.	
	($mil)	(%)	($mil)	(%)	($mil)	(%)	($mil)	(%)
Consumer Surplus								
Men	0.0	0.0	11.4	6.9	15.8	9.8	133.2	32.1
Women	0.0	0.0	25.7	5.4	49.3	10.9	40.8	10.6
Total	0.0	0.0	37.1	5.8	65.1	10.6	174.0	21.7
Revenue								
Men	0.0	0.0	22.2	10.1	30.6	14.4	83.5	40.7
Women	0.0	0.0	37.4	6.5	72.6	13.3	74.8	12.7
Total	0.0	0.0	59.6	7.4	103.2	13.6	158.2	20.0

Notes: Estimated gains to consumer surplus and firm revenue in millions of dollars and percentage. National NL model does not account for crowding via Ackerberg and Rysman (2005). Local NL results utilize tailored assortments.

Table 9: Consumer Welfare Gains with Price Changes

Online Retail Price Reduction	Consumer Welfare Increase ($mil)	(%)	Gains Due to Prices ($mil)	(%)
Baseline	37.1	5.8	—	—
1%	45.1	7.1	8.0	17.7
3%	61.1	9.9	24.0	39.3
5%	77.2	12.8	40.1	51.9
10%	117.4	21.0	80.3	68.4
15%	158.0	30.4	120.9	76.5
20%	198.8	41.5	161.7	81.3

Notes: Change in consumer welfare under a range of different price assumptions. Prior to online retail, consumers had access to fewer products and faced higher prices. With the entry of the online retailer, consumers gain access to the entire choice set and receive and across-the-board price reduction.

Table 10: Robustness: Overstatement of Consumer Welfare Increase

Assortment Size	Percent Increase			Absolute Increase ($ Millions)		
	Tailored	National	%Δ	Tailored	National	%Δ
Baseline	5.8	10.6	83.5	37.1	65.1	75.4
Threshold						
Mean Baseline	13.2	26.4	99.9	72.9	126.0	72.8
3000	57.6	107.3	86.5	226.1	319.2	41.2
6000	30.9	58.6	90.0	145.7	224.1	53.8
12000	13.0	26.1	100.1	72.0	124.6	73.2
24000	2.9	6.4	122.0	18.0	37.0	105.1

Notes: Results based on the Local RE parameter estimates in Table 5 and Table 6. The baseline assortment size is specified as the predicted values of $\log(a_\ell) = \beta_0 + \beta_1 \log(p_\ell) + \epsilon_\ell$, where a is the assortment size found in the Macy's and Payless data, and p is local population. The threshold assortment sizes impose the same assortment size in every local market.

Table 11: Robustness: Overstatement of Retail Revenue

Assortment Size	Percent Increase			Absolute Increase ($ Millions)		
	Tailored	National	%Δ	Tailored	National	%Δ
Baseline	7.4	13.6	83.2	59.6	103.2	73.2
Threshold						
Mean Baseline	17.1	34.0	98.7	115.4	194.7	68.7
3000	67.0	123.1	83.7	316.5	435.8	37.7
6000	37.4	70.4	88.2	213.9	320.6	49.9
12000	16.8	33.5	98.9	114.0	192.8	69.1
24000	4.2	9.5	123.8	32.8	66.6	103.4

Notes: Results based on the Local RE parameter estimates in Table 5 and Table 6. The baseline assortment size is specified as the predicted values of $\log(a_\ell) = \beta_0 + \beta_1 \log(p_\ell) + \epsilon_\ell$, where a is the assortment size found in the Macy's and Payless data, and p is local population. The threshold assortment sizes impose the same assortment size in every local market.