Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

11-1-2016

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

Thomas W. Quan

Kevin R. Williams

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Quan, Thomas W. and Williams, Kevin R., "Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail" (2016). *Cowles Foundation Discussion Papers*. 2512.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2512

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

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

By

Thomas W. Quan and Kevin R. Williams

November 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2054



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
Department of Economics
University of Georgia*

Kevin R. Williams
School of Management
Yale University†

November 2016‡

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 30% lower than previous studies.

JEL Classification: C13, L67, L81

*quan@uga.edu

†kevin.williams@yale.edu

‡We thank Pat Bajari, Steve Berry, Tat Chan, Judy Chevalier, Thomas Holmes, Kyoo il Kim, Aniko Oery, Amil Petrin, Fiona Scott-Morton, Ben Shiller, Mykhaylo Shkolnikov, Catherine Tucker, Joel Waldfogel, and especially Chris Conlon and Amit Gandhi for useful comments. We also thank the seminar and conference participants at 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, Harvard University, MIT, Boston College and the NBER IO Summer Institute. We also thank the Minnesota Supercomputing Institute (MSI) for providing computational resources and the online retailer for allowing us to collect the data.

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), 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 variety contribute to welfare?

The central idea of this paper is that the gains from online retail depend critically on the extent to which demand varies across geography and on how traditional brick-and-mortar retailers respond to local tastes.[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 want sandals – will benefit from the variety offered by online retail. Therefore, in order to quantify the gains from variety due to online retail, it is critical to estimate the extent to which demand varies both within and across locations.

This paper makes three contributions. Our first contribution is methodological. We augment the traditional nested logit demand model with across-market random effects. We propose it as a solution to the problem that, in big data sets such as ours, a large number of zero sales will arise with a large number of products and local demand. Given

[1] A large body of literature that has highlighted across-market differences in demand, including Waldfogel (2003, 2004, 2008, 2010), Bronnenberg, Dhar, and Dube (2009), Choi and Bell (2011), and Bronnenberg, Dube, and Gentzkow (2012). Crucially, Waldfogel (2008, 2010) also shows that the supply side responds to differences in tastes across geographic markets.

the sparsity of local sales, the underlying realizations of demand at each location cannot be identified. Our method allows us to focus on the distribution of demand across markets instead of its realizations. Second, it is well-known that commonly used discrete choice models may inflate the value of adding a large number of new products to the consumer's choice set; we demonstrate that our augmented model dampens this problem. Third, we provide estimates of the value of increased variety for a commonly purchased good, shoes. We use a novel data set from a large online retailer and show that abstracting from across-market demand heterogeneity leads to significantly overestimated gains from online variety.

Our estimation approach is necessitated by the characteristics of our data, which are shared by many highly disaggregated, large data sets. Demand estimation techniques, such as Berry (1994) and Berry, Levinsohn, and Pakes (1995), have been very successful in producing sensible estimates with aggregated data.[2] The maintained assumption is that as the size of the market increases, the sampling error in the observed market share, compared to the true underlying choice probability, approaches zero. However, with the proliferation of big data, researchers increasingly have access to very granular, high-frequency sales data. While fine granularity may contain additional information, it will often be the case that each type of shoe is not purchased in each market-period observation. Essentially, the purchase opportunities are rising as fast (or faster) than the number of purchases. This suggests assuming the market size is sufficiently large for the observed market share to be observed without sampling error is no longer reasonable.

In practice, observations with zero sales are often simply omitted from the analysis. This treats observed zeros as true zeros and assumes that there is no demand for these products. This approach 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.

[2] Aggregated across geographic markets, time, or products.

Second, the zeros are indicative of a small sample problem. This is particularly problematic for our setting because if uncorrected, we would overstate the degree of heterogeneity across markets (Ellison and Glaeser 1997) and understate the gains from increasing variety. For example, if we only observe one shoe sale for a particular market, it would suggest there are no gains from increasing variety because only one particular product is desired.

More recently, a number of potential solutions to the problem of zero sales have been employed. Within the generalized method of moments (GMM) framework,[3] proposed solutions include adjusting sales away from zero by making an asymptotically unbiased correction (Gandhi, Lu, and Shi 2014) or aggregating until the zeros disappear and adding micro moments to capture some disaggregated features of the data (Petrin 2002, Berry, Levinsohn, and Pakes 2004). With the severity of the local zeros problem in our application, the asymptotic correction has little effect because all local zeros are adjusted by the same amount (i.e. in Alaska, the unsold boot is adjusted to the same level as the unsold sandal) and estimated price elasticities remain too inelastic. Both of these factors lead to overstated welfare effects. Our approach to address local zeros is a form of the latter solution. However, unlike simple aggregation over products or geography, which would smooth over the heterogeneity we are interested in exploring, we are able to maintain narrow product definitions and retain information on local heterogeneity. We do so with the inclusion of across-market random effects that summarize the consumer heterogeneity important to the application at hand, but remains agnostic about its underlying sources.

To identify the random effects, we use micro moments derived from the fraction of zeros at the local level. Observed local zeros are rationalized by employing a finite sample multinomial, explicitly accounting for sampling. Our approach treats products with local zeros differently than the previous literature. For these products, our results lie in-between

[3]Another approach is to abandon the GMM framework in favor of maximum likelihood, such as in Chintagunta and Dube (2005). While there are trade-offs made when choosing between GMM and MLE. The two primary advantages of GMM are, first, product qualities can be estimated nonparametrically and, second, price endogeneity is addressed through exclusion restrictions/instrumental variables, rather than requiring a price model to be specified.

the extremes of dropping all of the zeros and adjusting all of the zeros by the same amount.

We also address the well-known econometric challenge that logit-style demand models tend to overstate welfare gains under large changes in the choice set. This occurs because each product in the choice set introduces a new dimension of unobserved consumer heterogeneity. This problem can be alleviated by flexibly modeling consumer heterogeneity with observables, e.g. Berry, Levinsohn, and Pakes (1995), Petrin (2002), Song (2007). Another approach, proposed by Ackerberg and Rysman (2005), is to introduce a crowding penalty that scales the variance of the logit error term. Our approach can be viewed in both lights. We model consumer heterogeneity across local markets using random effects, and we show that a function of the variances of our random effects corresponds to the Ackerberg and Rysman (2005) penalty term at the aggregate level. Whereas, in their implementation, they impose the ad-hoc assumption that each retail outlet sells only a select number products, we use micro data to estimate the penalty term. Our derivation provides a data-driven motivation for its use and estimation in applied work.

Using our model, we revisit the value of online variety. 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 of increasing variety to be seven to ten times larger than the competitive price effect. These gains have since been dubbed the "long-tail" benefit of online retail by Anderson (2004). These results have two major policy implications. First, the disproportionate impact of variety on welfare may suggest that antitrust enforcers should weigh changes in variety more than price effects. Second, it 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, suggesting online retail has led to a large 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, if past empirical methods have

overestimated the value of online variety, these implications may not hold.

To estimate the gains of increasing product variety, we use a detailed data set containing millions of geographically disaggregated footwear sales from a large online retailer. We show existing empirical approaches result in poor demand or welfare estimates because they either fail to address the sampling error at the local-level or they smooth over the heterogeneity of interest. Our model estimates confirm that demand varies greatly across markets. For example, our smallest estimate of across-market heterogeneity is for men's slippers, where we find a one standard deviation increase in the local demand shock of an average men's slipper is equivalent to a decline in price of $22. Due to this across-market heterogeneity, the existing literature overstates the value of products that mostly nobody buys because it fails to account for the fact that local assortments are tailored to local demand – a fact we confirm with new brick-and-mortar assortment data. When accounting for local heterogeneity, we find that consumer gains from increasing variety are over 30% lower than existing studies. Put another way, if local stores cater to the local demand, then the incremental value of online markets is much smaller because the average consumer already has access to many of the products he or she wants to purchase.

The rest of the paper is organized as follows. Section 2 presents the model and estimation procedure. Section 3 discusses our data and presents preliminary evidence of across-market heterogeneity. Results and counterfactuals are in section 4 and 5, respectively. Section 6 discusses the robustness of our findings, and the conclusion follows.

2 Model and Estimation

In this section, we first introduce the standard nested logit demand model, then show how we can augment the model with across-market random effects. Like Berry, Levinsohn, and Pakes (1995), our model adds random effects to the discrete choice setup. Whereas the usual random coefficients discrete choice setup allows for differences in consumer tastes across demographic groups, the random effect in our model enters as differences in

demand across locations. Since identifying the specific sources of heterogeneity is not our primary focus, we consolidate the random terms, which greatly reduces the computational burden. Further, while other aggregate models are consistent with across-market demand heterogeneity, it is difficult or impossible back out spatial information about this demand. Our approach allows us to investigate the distribution of local demand. We discuss the computational mechanics at the end of the section.

2.1 Standard Nested Logit Model

Each consumer solves a discrete choice utility maximization problem: Consumer i in location ℓ will purchase a product j if and only if the utility derived from product j is greater than the utility derived from any other product, $u_{i\ell j} \geq u_{i\ell j'}, \forall j' \in J \cup \{0\}$, where J denotes the choice set of the consumer and 0 denotes the option of not purchasing a product. We pursue a nested demand system where products can be grouped into mutually exclusive and exhaustive sets. Let c denote a nest, and note that every product j implicitly belongs to some nest c with the outside good belonging to its own nest.

To ease notation, we suppress the time script t. 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 $\zeta_{ic}+(1-\lambda)\varepsilon_{i\ell j}$ 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 only substitute to products within the same group and when $\lambda = 0$ the model collapses to the simple logit case.

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 characteristics, p_j is the price of product j, 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. This implies 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:

$$\begin{aligned}\pi_{\ell j} =& \pi_{\ell c} \cdot \pi_{\ell j/c} \\ =& \frac{\left(\sum_{j' \in c} \exp\{\delta_{\ell j'}/(1-\lambda)\}\right)^{1-\lambda}}{1 + \sum_{c' \in C}\left(\sum_{j' \in c'} \exp\{\delta_{\ell j'}/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{\exp\{\delta_{\ell j}/(1-\lambda)\}}{\sum_{j' \in c} \exp\{\delta_{\ell j'}/(1-\lambda)\}},\end{aligned} \tag{2.1}$$

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

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

$$\log(\pi_{\ell j}) - \log(\pi_{\ell 0}) = x_j\beta - \alpha p_j + \lambda \log(\pi_{\ell j|c}) + \xi_{\ell j}. \tag{2.2}$$

In the estimation of the standard model, the maintained assumption is that the size of each 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, high-frequency, highly disaggregated sales data is becoming increasingly available to researchers. Often, these data sets also contain large choice sets and with a large number of products, we may not expect to observe a sale for every product, especially at disaggregated market definitions. This suggests the market size assumption may no longer be reasonable. While aggregation over products or geography may be appealing in some settings, this would would smooth over the information contained in the disaggregated data. As we show below, our augmented nested logit model extracts information from both the aggregate and micro data.

2.2 Nested Logit Model Augmented with Random Effects

We propose a modification of the nested logit model that will allow us to aggregate over markets, while retaining information about across-market heterogeneity. To do so, we will need to place additional structure on the location-specific mean utilities. We assume 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, $\eta_{\ell j}$. Rearranging terms we have,

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

where δ_j is the mean utility of product j for the (national) population of consumers and $\eta_{\ell j}$ is a product-location-specific deviation. The heterogeneity in the random utility among consumers can then be decomposed into an "across-market" effect, $\eta_{\ell j}$, and a "within-market" effect, $\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}$. When $\eta_{\ell j} = 0$ for all $\ell \in L, j \in J$, the model reduces to the standard nested logit model from the previous section, where there is no distinction between local and national preferences.

Aggregating over location-specific choice probabilities across the distribution of locations yields the national choice probability

$$\pi_j = \int_\ell \pi_{\ell j} dF(\ell) = \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell j},$$

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

We could invert the market shares for each individual location ℓ, as in Equation 2.2, and proceed with linear instrumental variable methods to obtain estimates of the preference parameters. The local level residuals would then form estimates of $\xi_j + \eta_{\ell j}$. However, in many large data sets the sparsity of individual product sales within locations leads to selection bias. Adjusting local shares that are overwhelmingly zero using the asymptotic correction is also problematic. To circumvent these problems, we use a random-effects specification, where $\eta_{\ell j}$ is drawn independently from a normal distribution, $N(0, \sigma_j^2)$. Instead of attempting to recover each $\eta_{\ell j}$ directly, we estimate the variance of its distribution, σ_j^2.

Like Berry, Levinsohn, and Pakes (1995), our model corresponds to the addition of a random coefficient. However, unlike BLP, the random coefficient is constant for all consumers within a location and, importantly, we allow for sampling error in local level shares, which we discuss in the next subsection.[4] We allow for the possibility that prices are correlated with the national unobserved quality term ξ. This is appropriate when

[4]BLP introduces random coefficients through the interaction of product characteristics and consumer demographics. However, 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 severe small sample problem. To avoid the small sample problem at the local level, the BLP model could instead be estimated at the national level and preferences attributed to locations purely by their differences in observed demographics. However, to the extent that observable demographics fail to capture differences across markets, the degree of across-market heterogeneity will be understated, and hence, the gains from online variety will be overstated.

prices are set nationally. We assume prices are uncorrelated with η. We also maintain the usual assumption that all other observable product characteristics are exogenous with respect to the unobservables, ξ and η.

2.3 Integrating Over the Random Effects and the Market Share Inversion

To integrate out the across-market random effects, our proposed estimation exploits two important features our data, the large number of locations and the large number of products. Data sets containing these two features have become increasingly available and present the researcher with unique challenges for demand estimation.

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

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

Formally, we state this as a proposition.

Proposition 1. *For each product $j \in J$, applying the law of large numbers in L and integrating out over $\hat{\eta}$ gives*

$$\sum_{\ell \in L} \omega_\ell \pi_{\ell j}(\hat{\eta}_\ell; \delta, \lambda) - \pi_j \to_{a.s.} 0 \tag{2.3}$$

Proof. See Appendix A. ■

This suggests that, with a large number of locations, we can estimate the summation term over locations without knowing the exact realizations of η and thus, aggregated choice probabilities only depend on the variance of the across-market heterogeneity. Therefore, national demand can be expressed as

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

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

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

It is straightforward to show that the resulting inversion for our random effects model is[5]

$$\delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right). \tag{2.4}$$

Equation 2.4 relates δ_j to the aggregated share data, π_j, local population shares, ω_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 $\eta_{\ell j} = 0$, $\forall \ell \in L, j \in J$.[6] However, since $\eta_{\ell j}$ is an unknown random variable, unlike Berry (1994), we cannot simply recover mean utilities from observables.

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

$$\frac{1}{L}\sum_{\ell \in L} \omega_\ell \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\hat{\eta}_{\ell j}}{1-\lambda}\right\} - \frac{1}{L}\sum_{\ell \in L} E\left[\omega_\ell \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] \to_{a.s.} 0,$$

where, for each location ℓ, the expectation is over all products j. The complexity of this

[5] See Appendix A.

[6] 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

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

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

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

The conditional expectation not only depends on the local mean utilities of all other products, but the conditioning variable is the sum of lognormal random variables, which does not have a closed form expression for its distribution. Our setting involves a large number of products and we appeal to this fact to make further progress.

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. See Appendix A. ■

This proposition allows us to approximate the conditional expectation with the unconditional, which is simple to compute using the moment generating function of the normal distribution,[7] $E\left[\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right] = \exp\left\{\frac{1}{2}\frac{\sigma_j^2}{(1-\lambda)^2}\right\}$. Intuitively, when more products are added to a market the sum of random demand shocks is less informative about any individual shock. In the limit, knowing the sum of random shocks provides no information about an individual shock because high and low draws average out. We demonstrate with Monte Carlo exercises that using the unconditional expectation to approximate the conditional expectation performs well (see Appendix E).[8]

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

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

[8] We find in Monte Carlo exercises that the bias decreases quickly as the size of the choice set increases. For example, with 150 products and 200 locations the bias is upwards of 30%. If there are 525 products, the bias decreases to just 4-5% with 200 locations.

Finally, while 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 this assumption and given (σ, λ), we can then recover national mean utilities as function of observables $(\pi_j, \pi_{\ell c}, \pi_{\ell 0})$,

$$\delta_j = (1-\lambda)\left(\log\left(\pi_j\right) - \frac{1}{2}\frac{\sigma_j^2}{(1-\lambda)^2} - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\right)\right). \tag{2.5}$$

With a large number of products, the conditional expectation of our random effects converges to the unconditional, which does not depend on location. This has two important implications. First, δ_j can be recovered point-wise rather than requiring simultaneously solving a $J \times J$ system of equations for each location ℓ. This greatly reduces the computational burden of the problem, especially in situations with a large number of products. Second, at the national level, it suggests an adjustment that corresponds to the crowding penalty term proposed in Ackerberg and Rysman (2005). Define

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

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

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

Expanding this equation, we obtain[9]

$$\pi_j = \frac{\left(\sum_{j' \in c} R(\sigma_{j'}) \exp\{\delta_{j'}/(1-\lambda)\}\right)^{1-\lambda}}{1 + \sum_{c' \in C}\left(\sum_{j' \in c'} R(\sigma_{j'}) \exp\{\delta_{j'}/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{R(\sigma_j)\exp\{\delta_j/(1-\lambda)\}}{\sum_{j' \in c} R(\sigma_{j'})\exp\{\delta_{j'}/(1-\lambda)\}}. \tag{2.6}$$

That is, at the national level, the local random effects can be summarized as a function of the variances. Equation 2.6 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 is derived from an assumption on the number of retail outlets per product. Our penalty term is a summary statistic of the local preference unobservables. This allows us to motivate and identify the penalty term using micro data on across-market demand heterogeneity, as shown in the next subsection. However, both methods generate crowding through a similar channel: by lowering mean product utilities as the number of products increases.

The incorporation of random effects has important implications for our demand estimates. As more products enter the choice set, $R(\cdot)$ leads the product space to become more crowded. When modeling welfare this has the effect of diminishing the impact of each subsequent product entry. In our model, a large σ_j suggests the demand for product j is highly concentrated in particular geographic markets. The larger σ_j is, the smaller the mass of consumers with a high value for the product. Thus, the consumer welfare impact of removing products will tend to be smaller when σ_js are higher because fewer consumers are affected.

[9] 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 2.6.

2.4 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. Since 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.[10]

While 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 for our random effects. Let $P0_{\ell j}(\sigma; \delta, \lambda)$ be the probability that a product j has zero sales, given N_ℓ consumers are observed to make any purchase at location ℓ. We then define

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

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

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

where $s_{\ell j}$ is the observed location level market share for product j. Our micro moment

[10] 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.

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 proportion of locations may experience zero sales for a given product.[11]

2.5 Estimation Procedure

Having laid the foundation of our methodology, we turn to detailing the computational mechanics of the estimation. The model can be estimated using generalized method of moments (GMM). We start with the implementation of our 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 $\bar{\eta}_{\ell j}$ is an i.i.d. draw from a standard normal distribution. Given the assumptions on the individual level unobservable (GEV), there is a closed form expression for the location-product level choice probabilities, for any candidate value of σ and λ. We calculate the micro moments by conditioning on category. That is, for each product, we use the location level choice probability conditional on category, $\hat{\pi}_{\ell j|c}$, to simulate consumer purchases for each product at each location, holding the number of observed category purchases, $N_{\ell c}$,

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

fixed. In particular, the probability a product is observed to have zero sales at location ℓ is

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

i.e. the probability we observe $N_{\ell|c}$ sales within category c at location ℓ, none of which are good j.[12] 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.

With a candidate solution of σ and λ, the structure we have placed on the ηs allows us to integrate them out according to Equation 2.5 and recover national mean level 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 2.5,

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

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

[12] Alternatively, the micro moments can be formulated by conditioning on the inside shares or by taking the unconditional probability. The former is achieved by taking the inside sales as given and matching the probability of zero sales: $P0_{\ell j}(\sigma; \delta, \lambda) = (1 - \hat{\pi}_{\ell j|1})^{N_{\ell|1}}$, where $\hat{\pi}_{\ell j|1}$ is the probability of choosing good j conditional on making a purchase and $N_{\ell|1}$ is the number of purchases observed at location ℓ. The latter is achieved by matching the probability of zero sales using the unconditional choice probabilities: $P0_{\ell j}(\sigma; \delta, \lambda) = (1 - \hat{\pi}_{\ell j})^{N_\ell}$, where $\hat{\pi}_{\ell j}$ is the unconditional choice probability of product j and N_ℓ is the population of location ℓ. In Monte Carlo, we see no difference in the results based on the choice of formulation.

the within category share, but uncorrelated with the unobserved product quality.[13] The same instrument can be employed here, since $z_{j|c}$ 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 our moments and micro moments where $\theta = (\sigma, \lambda, \beta, \alpha)$ we have

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

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, $\theta^{(0)}$.[14] 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 $\widehat{\theta}^{(2)}$.

3 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 3.1). Next, we provide evidence of the localization of assortments using the brick-and-mortar assortment data (Section 3.2) and then demand-side across-market heterogeneity using the online

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

[14] 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. The estimation is performed using the Knitro solver with analytic gradients and takes up to two days to complete.

retail sales data (Section 3.3). 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 3.4).

3.1 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 a customer purchased more than a single pair of shoes, but we observe no other 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.[15] 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). While 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.[16] 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. While 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

[15]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.

[16]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.

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.[17] Since 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.

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% of shoes 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.[18] Unsurprisingly, we find that the stores with larger assortments tend to be located around larger population centers.

3.2 Localization of Brick-And-Mortar Retailers

Brick-and-mortar retailers are known for offering different product assortments across their networks.[19] 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

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

[18] 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.

[19] Ghemawat (1986) found 70% of Walmart's merchandise was common across stores, and 30% was tailored to local needs.

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."[20]

Of course, a firm's words may differ from their actions and while 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

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

This measure is bounded between zero and one. We use the minimum cardinality, rather than the cardinality of the union in the denominator, because we want this measure to capture differences in the composition of each store's inventory, not differences in assortment size. To further isolate differences in variety from differences in assortment size, we directly compare only locations of similar size. Figure 1 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.

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

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,[21] in addition to some locations where retailers maintain separate men's and women's stores. However, we do not believe there exists a substantial difference in relative costs across products that could lead to this geographic pattern since the vast majority of products are imported.

3.3 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 only be rationalized 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).[22] At the CSA+state-month level, we observe 213 local markets over 14 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

[21] 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.

[22] 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 New Jersey are contained in a CSA.

products is quite small across markets.[23]

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 2 plots the predicted values from a regression of a state's average annual temperature on the share of state revenue captured by boots and sandals. As expected, boots make up a greater share of revenue in colder states and a smaller share in warmer states. Conversely, the opposite relationship holds for sandals. This also suggests that consumers do not shop online just for products that are not available in traditional brick-and-mortar stores. For example, boots – rather than sandals – make up a sizable share of revenue in Alaska.

Other differences in demand across markets occur for less obvious reasons. In Figure 3, 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. While 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, while in parts of New York, New Jersey, and Massachusetts it makes up over 6% of sales. We will

[23] 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. For small cutoffs the difference in percent terms is very large but decreases as the cutoff increases between 3,000-5,000.

exploit this variation to help us identify across-market demand heterogeneity.

3.4 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 zip code-month level, 99.96% of products have zero sales. While simple aggregation over geography 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).

4 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. 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).[24] Our market sizes are proportional to the adult population for men and women, respectively.

Included in x are product ratings for comfort, look, and overall appeal and fixed effects for color, brand,[25] and time. The product ratings are time varying and reflect what the

[24] We find at finer levels of geography, such as zip code, the nearly 100% local zeros cause the micro moments to lose identifying power. We have confirmed this with Monte Carlo exercises, some of which appear in the Appendix. We choose CSA+state, compared to just CSA, since a large percentage of observed sales occur outside CSAs. For example, if we pursued the CSA market definition, we would drop all of sales to consumers in Alaska. Results dropping states are available on request.

[25] 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. In the estimation, we use the within transformation along the brand dimension to avoid explicitly estimating the large number of fixed effects.

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; J_b denote the set of products manufactured by brand $b \in B$; 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.

In principle, with our modeling assumptions and a large number of product-location observations, we could estimate 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 parametrize σ as a category-summer and category-winter random effect for boots and sandals, and as a category random effect for all other categories,[26]

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

Thus $mm(\cdot)$ contains $C + 2$ moments.[27] Overall, the estimation of the augmented nested logit model involves identifying up to fifteen parameters and the remaining mean utility parameters (α, β) using up to 360,000 observations.

We compare the estimates of our approach with a number of alternative models. For ease of exposition, we define these approaches now:

[26]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, while all other categories remain relatively stable over the course of the year.

[27]In addition to category, we have estimated the model using a parametric function of product rank, as well as interacting rank and category fixed effects. The results are similar to what we present here. We have noted more complicated functions, such as polynomials of rank interacted with category information are too computationally burdensome.

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 with aggregated (national) data

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 or "US." 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 or "AS." In our main discussion, only AS results are presented for Local RE and National NL for individual products. Category shares, which are strictly positive, are not adjusted.

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 empirical shares, with additional steps to minimize the asymptotic bias between the adjusted shares and the true conditional choice probabilities.[28]

4.1 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 - US; (2) Local NL-AS; (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

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

[28] A detailed discussion of the correction procedure can be found in Appendix C.

and for local level estimates, price elasticities are computed as

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

Specifications (1) and (2), Local NL-US and Local NL-AS, illustrate the selection bias generated by the severity of the zeros problem (95% of observations), even when employing adjusted shares. Of particular concern for us are the price coefficients and the nesting parameters. In each case, one or both of these parameters are biased toward zero. In three of the four instances the estimates are biased past zero. Using unadjusted shares, we find the nesting parameter is negative for women and the price coefficient becomes positive for both men and women. The impact of the selection bias in the price and nesting parameters imply price elasticities that are inelastic using unadjusted shares and positive using adjusted shares. Elasticity estimates using unadjusted local shares for men and women's 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.[29]

In the last two columns, we report the results from estimating the National NL and Local RE models. All coefficients have the correct sign and are significant. They are also similar in magnitudes which is unsurprising – the two models predict the same aggregate mean utilities. These two approaches yield similar average product level price elasticities, (-3.6, -3.3) and (-3.0, -2.1) for men and women, respectively.[30] The nesting parameter for men's shoes, (0.81, 0.57), suggests substitution within category is important. For women's shoes, we obtain nesting parameters of 0.37 and 0.28, implying lower substitutability amongst shoes within category, compared to men. Despite the similarity in mean demand

[29]Overall, traditional approaches using adjusted and unadjusted shares perform poorly. In Appendix D, we examine alternative specifications using unadjusted and adjusted share. For example, we estimate models market-by-market and show the results are similarly poor. This suggests simply adding local flexibility to the demand model does not improve the estimates because the primary issues stem from the zeros problem, not lack of specification flexibility.

[30]The empirical literature on shoe demand is limited. Roberts, Xu, Fan, and Zhang (2012) 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.

parameters, the Local RE model has two primary advantages over the National NL model. First, it retains information on the distribution of heterogeneity across locations, allowing us to investigate the premise of this paper. The importance of this distinction will be highlighted in the following section when we perform counterfactual analyses. Second, as highlighted by Ackerberg and Rysman (2005), standard nested logit models will overstate the size of the value of increasing variety and the Local RE model addresses this concern.

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. Look and comfort are all positive, but have smaller effects than comfort. Meanwhile, our indicator for no reviews takes on positive signs for both men's and women's shoes. This variable largely captures the demand for new products before there has been an opportunity to review them. New products often benefit from additional promotion and advertising, and it is likely that the positive effect of having no review actually reflects the additional promotion, rather than a desire to purchase shoes that have not been reviewed.

4.2 Across-Market Heterogeneity

A key advantage of the Local RE model for our application is that it rationalizes the distribution of local demand and provides us with estimates of across-market demand heterogeneity. Our estimates for the across-market demand heterogeneity in the Local RE model are presented in Table 7.

We find all the across-market heterogeneity parameters to be statistically significant. More importantly, these parameters are highly significant economically. For example, the smallest statistically significant σ_c for men's and women's shoes is slippers at 0.29 and 0.46, respectively. To put these numbers in perspective, a one standard deviation increase in a slipper's draw of $\eta_{\ell j}$ is equivalent to a decline in price of around \$22 for men and \$38 for women. These large effects will have important implications for consumer welfare analysis, as we will see in the following section. Finally, while 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 \$22-31 for men and \$38-50 for women, depending on category.

Across-market demand heterogeneity is important for rationalizing the distribution of local sales in the data. Figure 4 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 since 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 85,622 men's and 65,934 women's product-location pairs that are observed in the data to be zero.

5 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 5.1). We compare consumer surplus and retail revenue under the large (observed) 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 are tailored to local demand and (2) where local assortments are standardized, which is analogous to the counterfactuals found in the existing literature. 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 5.2).

5.1 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. 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. While 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. Since 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

$$\log(a_\ell) = \beta_0 + \beta_1 \log(\text{pop}_\ell) + \varepsilon_\ell,$$

where a_ℓ is the number of unique varieties from the exercise above and pop_ℓ is 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

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

and retail revenue is defined as

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

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 specifications. Our estimates for the gains of increased 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 $52.3 million, or 8.3%. Our interpretation is that these numbers are sizable, but are about 30% lower than the gains without tailored assortments (Local RE - National Assortment), which is the exercise performed in the existing literature. We find the overstatement in consumer welfare to be over 35% in absolute terms and over 40% 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. For example, if the national ranking highly rate sneakers and sandals, there will be too few boots for consumers in Alaska.

The Local NL provide vastly different estimates for the gains of 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 satisfies the entirety of local consumer demand and we estimate the consumer welfare gains to be nonexistent. Further, using the adjusted shares results in nonsensical estimates of consumer surplus, due to a positive price coefficient.

Finally, by comparing our Local RE and the National NL model (last column), we can see the effect of failing to account for the variance of the logit error in the vain 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 almost triple that with our estimator and 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 limits 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 4.2% higher under the latter. This suggests that there may be a significant incentive for brick-and-mortar stores to cater to local demand, depending on the potential dis-economies of scale due to localization.

5.2 Long Tail Analysis

Our results have important implications for the understanding of the long-tail phenomenon observed in online retail. Our data suggests that "shorter" revenue tails at the local level underlie the long tail at the national level. Using the raw sales data, Figure 5 illustrates how local level "short" tails can aggregate to a national level long tail. It plots the cumulative share of revenue going to the top K products (x-axis) for the following scenarios: median market (by number of monthly sales), middle 10% (p45-p55), middle 50% (p25-p75), and all the data.

For the median market, we can see that there is an extremely short tail, with fewer than 2,000 products making up total local revenue. The next line (p45-p55) aggregates the sales data for the middle 10% of markets. Since 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. Sales become

less concentrated among the top products producing a lengthening effect of the revenue tail. Finally, using the middle 50% of markets (p25-p75) shows the lengthening effect of the tail as more markets are combined. Hence, the plow shows that simply aggregating over markets creates a long tail, even though each individual market demands far fewer varieties of shoes.

On the other hand, our data and many large data sets suffer from a small samples problem at the local level. The raw data suggests an incredibly short tail at the local level – which translated to zero welfare gains by using empirical local shares in the welfare analysis. We can correct for sampling in this long tail analysis by utilizing the results of the Local RE model and simulating a large number of purchases for each of our local markets. Figure 6 contains the same median market (data) revenue curve along with the national revenue curve found in Figure 5. We add a line called "Median (Simulated)" which removes the small sample problem for that location. There are two important results. First, it suggests local tails are quite a bit longer than suggested by the raw data. This value is captured in the welfare analysis by how the Local RE model treats zeros. The second finding is that the tail is much shorter than the national level curve, which is consistent with the story of across market demand heterogeneity.

6 Robustness and Additional Insights

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

6.1 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 gain from online variety. Figure 7 shows this relationship. The plot is centered on the estimated σ (y-axis) and corresponding percent of local zero sales and estimated gain in welfare (x-axis) of the Local RE model. As across-market heterogeneity increases, corresponding to σ between 100% and 200%, the proportion of zeros increases. The model is identified by choosing σ to match this feature of the data. The second feature highlighted in the figure is that as across-market demand heterogeneity increases, the gains of online variety decrease. This is because each location has stronger preferences for a smaller subset of products. Since local retailers cater to these preferences, the additional value created by the large online choice set is smaller. The relationship is the exact opposite for σ between 0% and 100%.

6.2 Welfare and Counterfactual Choice Sets

Our results are based on a specified counterfactual choice set. Here we conduct sensitivity analysis to the choice set size. We find 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 9 presents the change in consumer welfare and the size of the overstatement resulting from various thresholds of the counterfactual choice set, respectively. For comparison, we also include our baseline results from the previous section 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 8, 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 (dash).

Table 10 presents the 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 16% 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 9 graphs the increase in retail revenue due to local customization of assortments, measured in millions of dollars (solid) and as a percentage (dash). 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.

6.3 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 10 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 lines). To test how sampling impacts the revenue curve, we remove all products in which only a single local sale occurs (dashed lines).

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.

7 Conclusion

In this paper, 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 confront the severe small sample problem in our data, while retaining information on the across-market heterogeneity of interest to us. Our estimates suggest 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 consumer welfare and firm strategy. 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 16% additional revenue by localizing assortments. For consumers, our calculations suggest that abstracting from across-market demand heterogeneity overstates the gain in consumer welfare from online variety by about 35-40%.

There are several potential avenues through which online retail benefits consumers. For example, the entry of online firms creates competition with local brick-and-mortar retailers. This can lead to a reduction in prices and an increase in consumer welfare. The variety channel is another avenue through which online retail can increase consumer welfare: the large online choice set allows for better product matching compared to the limited selection available at physical stores. Our results suggest that this channel may be

substantially less important and that the long tail of online retail may contain substantially less value than previously thought because local preferences are correlated.

Although we bring in new, rich data and propose new methodology to estimate demand with 95% local zeros, both the data and methodology have limitations. With our data, we have to abstract from consumer search. Like the existing literature, we lack data on pre-internet assortments and resort to counterfactual exercises that assume the stocking decisions of brick-and-mortar retailers have been unaffected by the advent of online retail. Additionally, we assume that brick-and-mortar retailers are able to perfectly predict consumer demand.[31] However, as long as there was some degree of local customization in brick-and-mortar assortments before the internet, our main conclusion holds: it is important to account for across-market demand heterogeneity when estimating the gain from online variety. The main drawbacks of the methodology are it only allows us to estimate the distribution of local demand and requires national pricing if endogeneity is a concern. A potentially interesting area of future research is addressing sampling in more flexible demand systems incorporating location-level or individual-level data.

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 (2015): "Quality Predictability and the Welfare Benefits from New Products: Evidence from the Digitization of Recorded Music," Working paper.

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

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

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," Econometrica, 63(4), 841–890.

[31] Aguiar and Waldfogel (2015) explore the benefit of having a long tail of products when quality is not perfectly predictable.

——— (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.

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

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

Brynjolfsson, E., Y. J. Hu, and M. D. Smith (2003): "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," Management Science, 49(11), 1580–1596.

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.

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

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.

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.

Ghemawat, P. (1986): Wal-Mart stores' discount operations. Harvard Business School Boston.

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.

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. Y. Xu, X. Fan, and S. Zhang (2012): "A structural model of demand, cost, and export market selection for chinese footwear producers," Working paper.

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

Song, M. (2007): "Measuring Consumer Welfare in the CPU Market: An application of the pure-characteristics demand model," The RAND Journal of Economics, 38(2), 429–446.

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.

A Proofs

Proposition 1. *For each product $j \in J$, applying the law of large numbers in L and integrating out over η gives*

$$\sum_{\ell \in L} \omega_\ell \pi_j \left(\eta_\ell; \delta, \lambda\right) - \pi_j \to_{a.s.} 0 \tag{A.1}$$

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

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

Aggregating over local choice probabilities gives

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

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

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

Plugging this equation for $D_{\ell c}$ into the denominator of the aggregated choice probabilities, gives

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

We now apply Kolmogorov's Strong Law of Large Numbers to show that the sum of local choice probabilities converges almost surely to the sum of the expectations of the local choice probabilities. That is,

$$\sum_{\ell \in L} \omega_\ell \pi_j \left(\eta_\ell; \delta, \lambda\right) - \sum_{\ell \in L} \omega_\ell E\left[\pi_j \left(\eta_\ell; \delta, \lambda\right)\right] \to_{a.s.} 0.$$

Note that $\pi_j \left(\eta_\ell; \delta, \lambda\right)$ 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,

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

We break down this stage of the proof into several components that culminate in the application of Kolmogorov's Strong Law of Large Numbers.

- We first show that $\left(\frac{1}{\pi_{\ell c}}\right)^K$ has finite mean and variance, for all $K \geq 0$.
- It is then easy to show that $\left(\omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)$ has finite mean and variance.
- Finally, we show that the conditions needed to apply Kolmogorov's SLLN hold.

From the structure of the nested logit demand model,

$$\left(\frac{1}{\pi_{\ell c}}\right)^K = \left(\frac{\sum_{c'} D_{\ell c'}^{1-\lambda}}{D_{\ell c}^{1-\lambda}}\right)^K$$

for $K \geq 0$ and $\lambda \in (0,1)$. Since $f(x) = x^K$ is a monotonic transformation of x, we have that

$$\left(\frac{\sum_{c'} D_{\ell c'}^{1-\lambda}}{D_{\ell c}^{1-\lambda}}\right)^K \leq \left(\frac{|C|^\lambda \left[\sum_{c'} D_{\ell c'}\right]^{1-\lambda}}{D_{\ell c}^{1-\lambda}}\right)^K,$$

where $|C|$ denotes the cardinality of C, i.e. the number of categories including the outside good. This step is achieved by applying Jensen's Inequality of the form $\{a_i\}_{i=1}^J : a_i > 0, \forall i$

$$(a_1 + ... + a_J)^k \leq J^{k-1}(a_1^k + ... + a_J^k) \text{ if and only if } k \geq 1,$$

where the the inequality is reversed if $k \leq 1$ (as used in the first inequality). Let $m = (|C|^\lambda)^K$, we have

$$\left(\frac{|C|^\lambda \left[\sum_{c'} D_{\ell c'}\right]^{1-\lambda}}{D_{\ell c}^{1-\lambda}}\right)^K = m\left(\left[\frac{\sum_{c'} D_{\ell c'}}{D_{\ell c}}\right]^{1-\lambda}\right)^K = m\left(\frac{\sum_{c'} D_{\ell c'}}{D_{\ell c}}\right)^{\tilde{K}},$$

where we let $\tilde{K} = K(1-\lambda)$. The numerator is equal to $\sum_{j \in J} \exp\left(\frac{\delta_j + \eta_{\ell j}}{1-\lambda}\right)$, i.e. summed over all products, and the denominator is similarly defined but summed over only $j \in c$. Define $\tilde{N}_j = \frac{\delta_j + \eta_{\ell j}}{1-\lambda}$ and let $J_c/J > \alpha > 0$ for each category. Then we have

$$m\left(\frac{\sum_{c'} D_{\ell c'}}{D_{\ell c}}\right)^{\tilde{K}} = m\left(\frac{\sum_{j \in J} \exp\left(\tilde{N}_j\right)}{\sum_{j' \in c} \exp\left(\tilde{N}_{j'}\right)}\right)^{\tilde{K}} \leq m\left(\frac{\frac{1}{J}\sum_{j \in J} \exp\left(\tilde{N}_j\right)}{\frac{J_c}{J}\frac{1}{J_c}\sum_{j' \in c} \exp\left(\tilde{N}_{j'}\right)}\right)^{\tilde{K}} \leq m\left(\frac{\frac{1}{J}\sum_{j \in J} \exp\left(\tilde{N}_j\right)}{\alpha \exp\left\{\tilde{N}_1\right\}}\right)^{\tilde{K}},$$

where arbitrary element 1 is such that $\exp\left\{\tilde{N}_1\right\} = \min_{j' \in c} \exp\left(\tilde{N}_{j'}\right)$. Hence this element is by construction less than the average, and we can define such element for each inside category share.

Suppose $\tilde{K} \leq 1$. Then by Jensen's Inequality,

$$mE\left[\left(\frac{1}{J}\sum_{j \in J} \frac{\exp\left(\tilde{N}_j\right)}{\alpha \exp\left(\tilde{N}_1\right)}\right)^{\tilde{K}}\right] \leq m\left(E\left[\frac{1}{J}\sum_{j \in J} \frac{\exp\left(\tilde{N}_j\right)}{\alpha \exp\left(\tilde{N}_1\right)}\right]\right)^{\tilde{K}}.$$

We have

$$E\left[\frac{1}{J}\sum_{j\in J}\frac{\exp\left(\tilde{N}_j\right)}{\alpha\exp\left(\tilde{N}_1\right)}\right]=E\left[\frac{1}{J}\sum_{j\in J}\frac{\exp\left(\tilde{N}_j-\tilde{N}_1\right)}{\alpha}\right]=\frac{1}{J}\sum_{j\in J}E\left[\frac{\exp\left(\tilde{N}_j-\tilde{N}_1\right)}{\alpha}\right].$$

This is a finite sum of the expectations of random variables, each with finite mean and variance. Hence,

$$E\left[\left(\frac{1}{\pi_{\ell c}}\right)^K\right]\leq m\left(\frac{1}{J}\sum_{j\in J}E\left[\frac{\exp\left(\tilde{N}_j-\tilde{N}_1\right)}{\alpha}\right]\right)^{\tilde{K}}<\infty.$$

Next, suppose $\tilde{K}>1$. By Jensen's Inequality, we have

$$E\left[\left(\frac{1}{J}\sum_{j\in J}\frac{\exp\left(\tilde{N}_j\right)}{\alpha\exp\left(\tilde{N}_1\right)}\right)^{\tilde{K}}\right]\leq E\left[J^{\tilde{K}-1}\frac{1}{J^{\tilde{K}}}\sum_{j\in J}\left(\frac{\exp\left(\tilde{N}_j\right)}{\alpha\exp\left(\tilde{N}_1\right)}\right)^{\tilde{K}}\right].$$

The right hand side is equal to

$$J^{-1}\sum_{j\in J}E\left[\frac{\exp\left(\tilde{K}\tilde{N}_j-\tilde{K}\tilde{N}_1\right)}{\alpha}\right],$$

which, again, is a finite sum of expectations of random variables, each with finite mean and variance. Hence,

$$E\left[\left(\frac{1}{\pi_{\ell c}}\right)^K\right]\leq J^{-1}\sum_{j\in J}E\left[\frac{\exp\left(\tilde{K}\tilde{N}_j-\tilde{K}\tilde{N}_1\right)}{\alpha}\right]<\infty.$$

The variance result, then follows

$$V\left[\left(\frac{1}{\pi_{\ell c}}\right)^K\right]=E\left[\left(\frac{1}{\pi_{\ell c}}\right)^{2K}\right]-\left(E\left[\left(\frac{1}{\pi_{\ell c}}\right)^K\right]\right)^2<\infty.$$

by applying the expectation result to both terms.

Next, by applying the result above, since $\omega_\ell\pi_{\ell c}\left(\pi_{\ell 0}\right)^{\frac{1}{1-\lambda}}$ is a random variable bounded between 0 and 1 and $\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}$ is a lognormal random variable, it is the straightforward to show that $\left(\omega_\ell\pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)$ has finite mean and variance. Note also that their expectations only differ in ω_ℓ, which is bounded between 0 and 1. Thus, the variance is bounded above when $\omega_\ell=1$ and

$$\sum_{\ell=1}^{\infty}\frac{V\left[\omega_\ell\pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right]}{\ell^2}<\sum_{\ell=1}^{\infty}\frac{V\left[\pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right]}{\ell^2}<\infty.$$

Therefore, by Kolmogorov's Strong Law of Large Numbers

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

Since δ_j is a constant, rearranging it back into the expectation obtains our result.■

Corollary: Suppose $J_c/J > \alpha > 0$, for all J. Assume sequence $\{\delta_j\}_{j=1}^J$ is bounded for all J and the variance of η is bounded for all J. Then, Proposition 1 holds as $J \to \infty$.

Proof. Proposition 1 holds as $J \to \infty$. Given the sequence of $\{\delta_j\}_{j=1}^J$ is bounded, and the variances of η are finite, then $E\left[\exp\left(\tilde{K}\tilde{N}_j - \tilde{K}\tilde{N}_1\right)\right]$ exists and is finite using the moment generating function of normal distributions. The rest of the proof follows. ■

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 $\pi_{\ell c}$.

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

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

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

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

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

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:

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

Proof. By the definition of variance,

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

By Jensen's Inequality,

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

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. $\frac{1}{I_c} D_{\ell c} \to_d \delta^c$ for each c.[32] ■

Thus $E\Big[\Psi_\ell(J;c)\Big|\bar{\pi}_{\ell c}\Big]$ converges to a constant and $\bar{\pi}_{\ell c} \cdot E\Big[\Psi_\ell(J;c)\Big|\bar{\pi}_{\ell c}\Big]$ converges to a constant by the law of large numbers. By an analogous argument to the lemma, $E\Big[\bar{\pi}_{\ell c} \cdot E\Big[\Psi_\ell(J;c)\Big|\bar{\pi}_{\ell c}\Big] \,\Big|\, \pi_{\ell 0}\Big]$ 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,

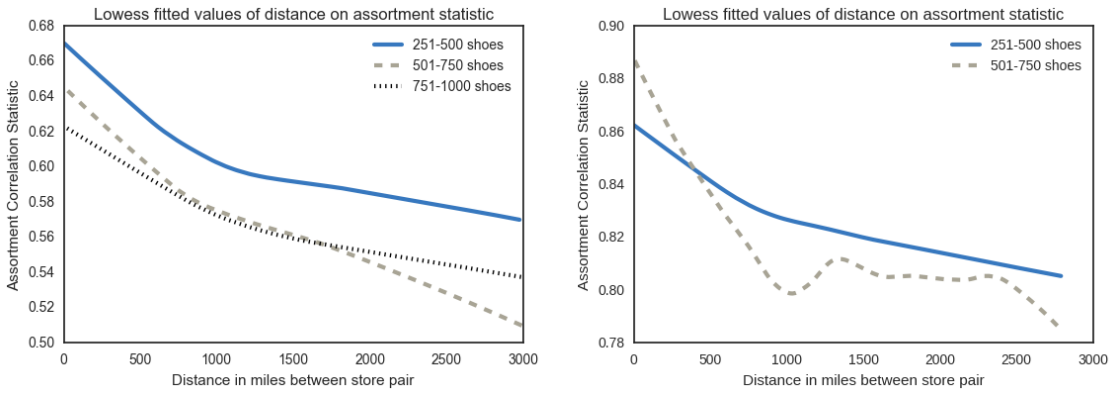
$$E\Big[\exp\Big\{\frac{\eta_{\ell j}}{1-\lambda}\Big\} \,\Big|\, \pi_{\ell c}, \pi_{\ell 0}\Big] \to_d E\Big[\exp\Big\{\frac{\eta_{\ell j}}{1-\lambda}\Big\}\Big] \quad \text{as } J \to \infty.$$

■

[32] 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.

B Tables and Figures

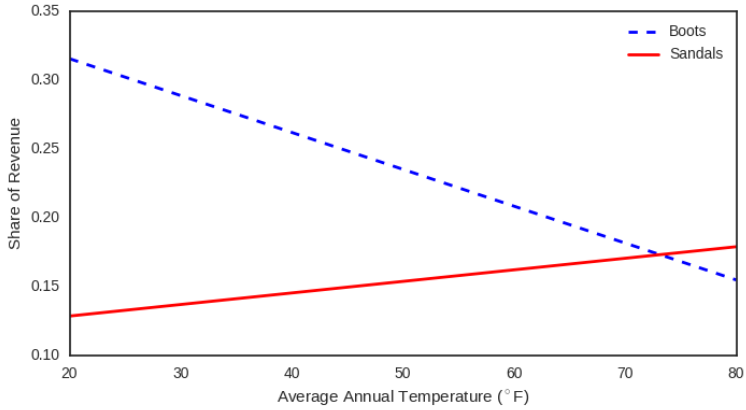
Figure 1: Assortment Overlap by Distance



(a) Macy's (b) Payless

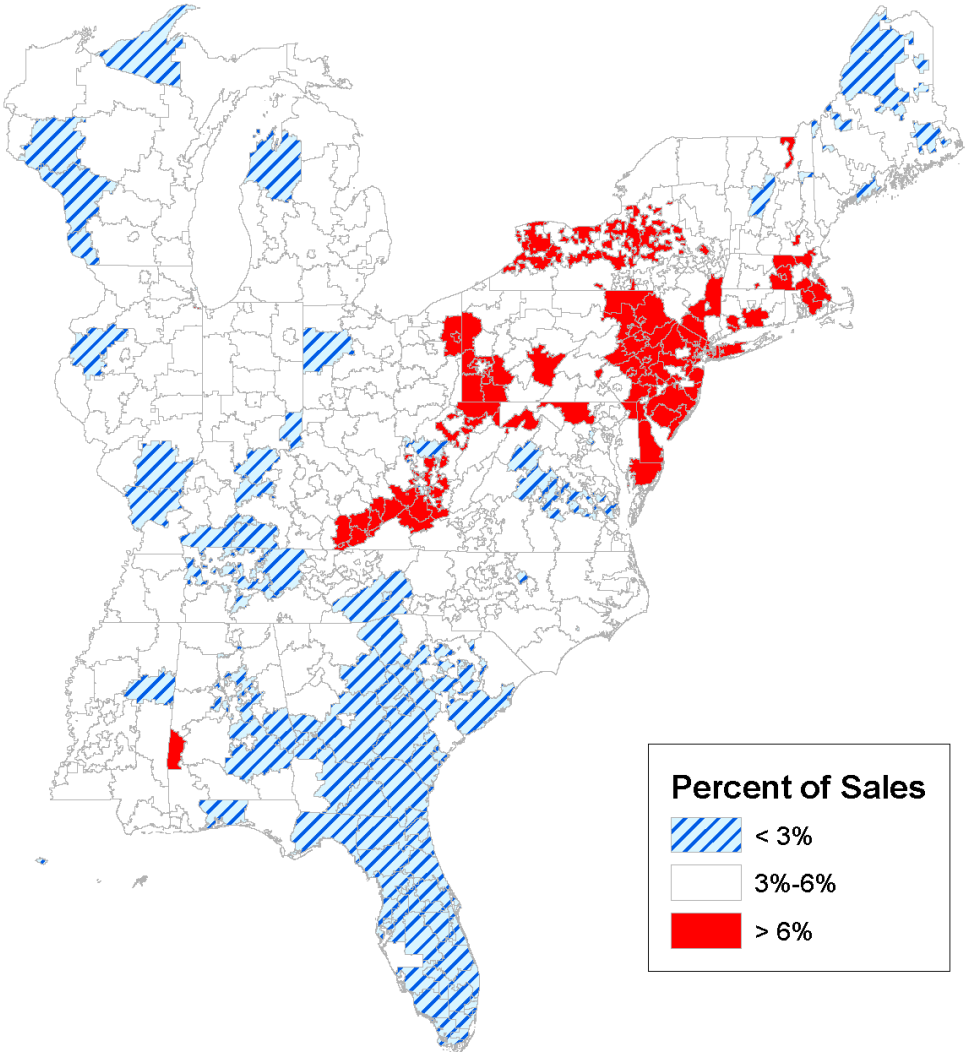
Note: Lowess fitted values of assortment overlap across stores in the network. Analysis split across stores with similar assortment sizes.

Figure 2: Boots vs. Sandals: Revenue by Temperature



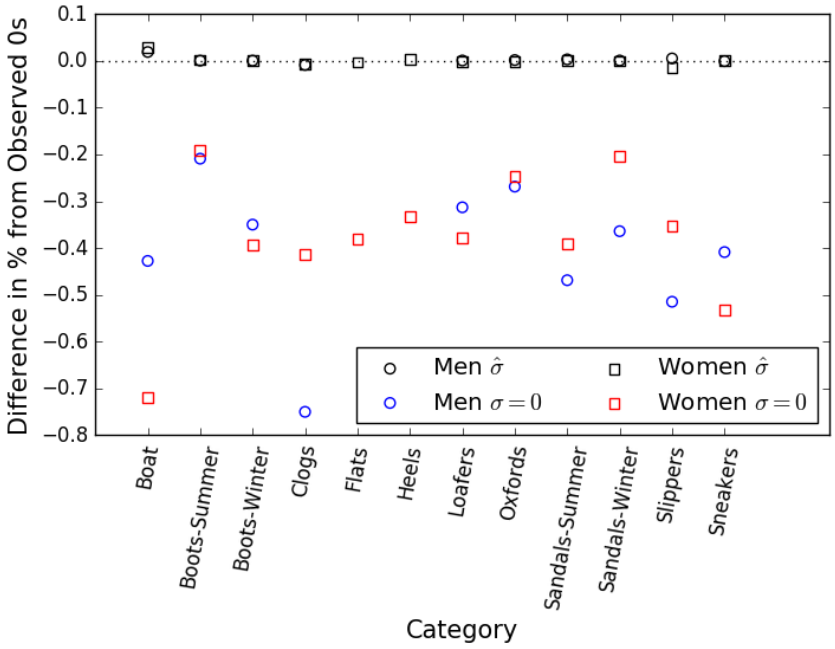
Note: Fitted values from a linear regression of average annual state temperatures on the sales of boots and sandals as a share of state revenue.

Figure 3: Revenue Share of a Popular Brand Across Zip3s



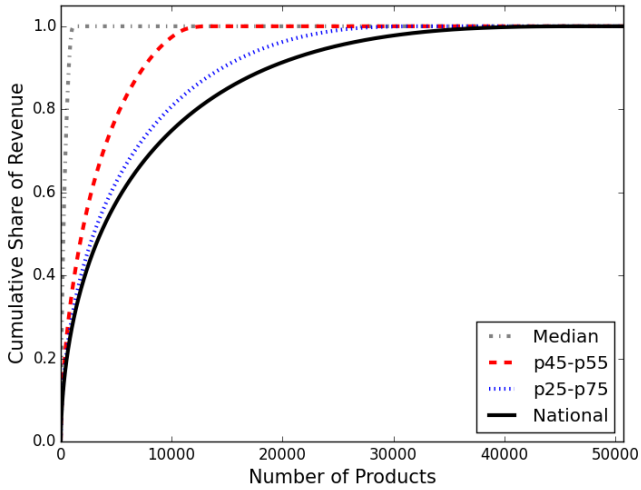
Note: Map of Eastern US Zip3s – the first 3 digits of a 5-level zip code. The color of the Zip3 corresponds to the local revenue share of a popular brand in the data set. Sales of the brand are concentrated in the Eastern US.

Figure 4: Predicted Zeros Without Across-Market Demand Heterogeneity



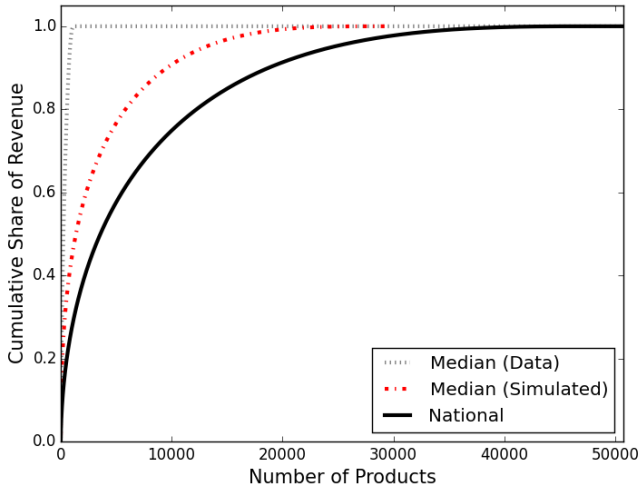
Note: For each product category, the difference between the observed percentage of zeros and the predicted percentage of zeros. Predicted zeros come from simulation of sales using the estimated level of across-market demand heterogeneity, $\hat{\sigma}$, and assuming no across-market demand heterogeneity,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

Figure 5: Aggregating to the Long Tail



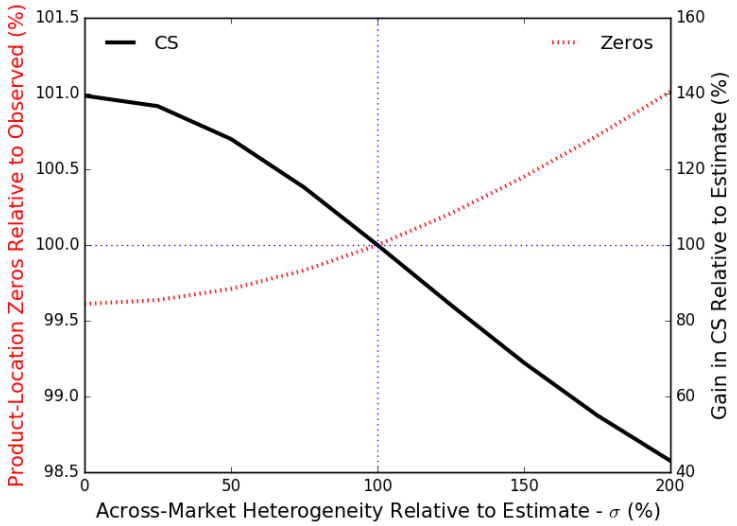
Note: For varying levels of aggregation, the cumulative share of revenue going to the top products.

Figure 6: Local Tail: Correcting for Small Samples



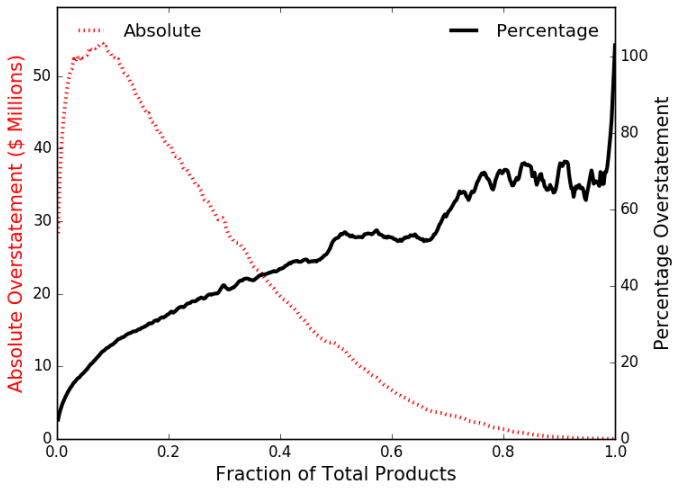
Note: For the median local market (CSA+state, by number of monthly sales), the cumulative share of revenue going to the top products, as seen in the data (dot) and simulated using our estimated demand system (dash-dot). For comparison, we also include the national level revenue distribution (solid).

Figure 7: Impact of Across-Market Heterogeneity on Zeros and Welfare



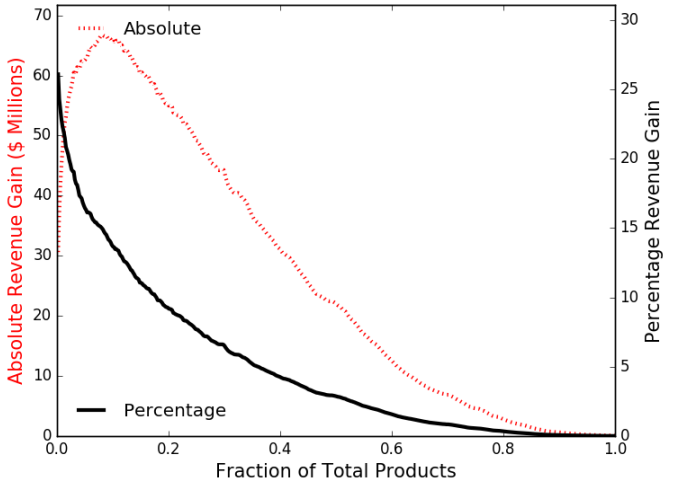
Note: The predicted number of product-location zeros and estimated consumer welfare for different levels of σ_c. Along the x-axis, "0" indicates no across-market demand heterogeneity, "100" corresponds to our estimates, and "200" corresponds with two times our estimated σ_c.

Figure 8: Overestimation of Consumer Welfare Gains



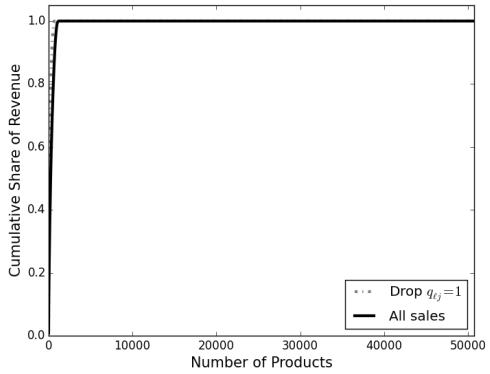
Note: The overstatement in consumer surplus gains, by counterfactual assortment size, when assuming a nationally standardized assortment vs. a locally customized assortment measured in dollars (red, dotted) and percentage (black, solid).

Figure 9: Increase in Retail Revenue from Localized Assortments

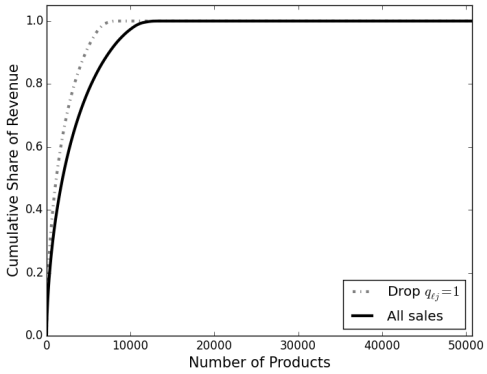


Note: The gain in retail revenue, by local retailer assortment size, when moving from a nationally standardized assortment to a locally customized assortment measured in dollars (red, dotted) and percentage (black, solid).

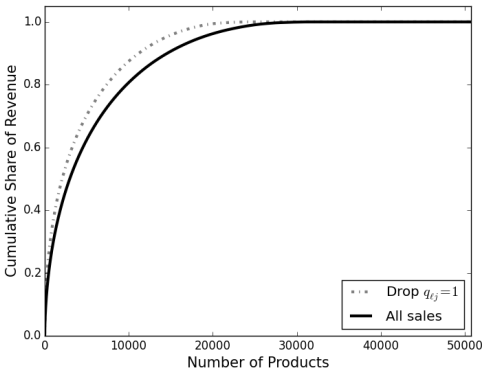
Figure 10: Demand Aggregation Dropping Single Sale Observations.



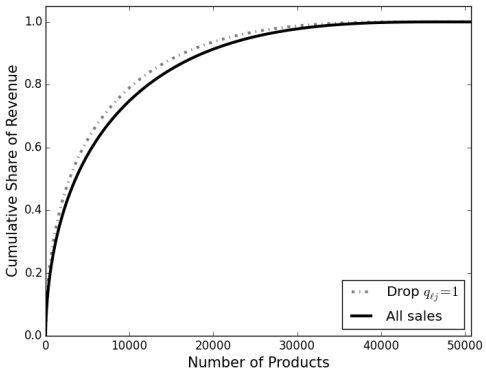
(a) Median Market



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



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



(d) Aggregation of All Markets (National)

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

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. $N_{\text{Macy's}} = 93,602,700$, $N_{\text{Payless}} = 69,451,866$.

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 - Rejection 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
Combined Statistical Area (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	–

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 dropped as all sales in these states are assigned to 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
Combined Statistical Area (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

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 of Zero Sales	Market Top 500	
		Market	National
SKU (shoe + style)	95.54	67.05	7.23
Shoe	93.10	73.39	19.07
		Market Top 10	
		Market	National
SKU (shoe + style)	95.54	7.59	0.50
Shoe	93.10	9.10	2.18
Brand	59.27	33.91	25.48

Time horizon fixed at the monthly level and geography 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^{***}	-0.013^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
Comfort	0.033^{***}	-0.003^{***}	0.002	0.001
	(0.003)	(0.000)	(0.003)	(0.007)
Look	0.000	-0.003^{***}	0.007	0.018^{*}
	(0.003)	(0.000)	(0.004)	(0.009)
Overall	0.045^{***}	0.000^{*}	0.059^{***}	0.133^{***}
	(0.003)	(0.000)	(0.004)	(0.008)
No Review	0.266^{***}	-0.041^{***}	0.274^{***}	0.598^{***}
	(0.012)	(0.000)	(0.022)	(0.043)
λ	0.103^{***}	0.989^{***}	0.814^{***}	0.570^{***}
	(0.010)	(0.000)	(0.006)	(0.006)
σ	—	—	—	*
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
Price Elast.				
Mean	−0.876	10.996	−3.635	−3.384
Std. Dev.	(0.619)	(8.277)	(2.735)	(2.546)

Notes: Estimated at the monthly level. "Local NL" (1) estimates nested logit demand at the CSA-State level with adjusted 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^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
Comfort	0.027^{***}	-0.008^{***}	0.045^{***}	0.050^{***}
	(0.002)	(0.000)	(0.007)	(0.007)
Look	0.006^{***}	0.003^{***}	0.025^{***}	0.019^{**}
	(0.002)	(0.000)	(0.008)	(0.008)
Overall	0.047^{***}	-0.012^{***}	0.134^{***}	0.148^{***}
	(0.002)	(0.000)	(0.009)	(0.008)
No Review	0.263^{***}	-0.158^{***}	0.781^{***}	0.783^{***}
	(0.009)	(0.001)	(0.032)	(0.031)
λ	-0.034^{***}	0.932^{***}	0.370^{***}	0.280^{***}
	(0.003)	(0.000)	(0.010)	(0.010)
σ	—	—	—	*
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
Price Elast.				
Mean	−0.592	5.644	−3.055	−2.127
Std. Dev.	(0.474)	(4.939)	(2.672)	(1.861)

Notes: Estimated at the monthly level. "Local NL" (1) estimates nested logit demand at the CSA-State level with adjusted 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	0.323***	0.595***
	(0.008)	(0.014)
Boots - Summer	0.404***	0.593***
	(0.008)	(0.010)
Boots - Winter	0.370***	0.526***
	(0.007)	(0.008)
Clogs	0.309***	0.541***
	(0.008)	(0.009)
Flats	–	0.548***
		(0.009)
Heels	–	0.545***
		(0.008)
Loafers	0.328***	0.567***
	(0.006)	(0.011)
Oxfords	0.317***	0.539***
	(0.006)	(0.016)
Sandals - Summer	0.317***	0.516***
	(0.006)	(0.008)
Sandals - Winter	0.349***	0.558***
	(0.009)	(0.010)
Slippers	0.292***	0.464***
	(0.009)	(0.013)
Sneakers	0.313***	0.545***
	(0.005)	(0.008)

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	Adjusted Shares	Tailored Assort.	National Assort.	National Assort.
Consumer Surplus						
Men	$mil	0.0	-2342.7	16.0	21.8	133.2
	%	0.0	41.2	9.1	12.8	32.1
Women	$mil	0.0	-1710.2	36.3	49.5	40.8
	%	0.0	45.7	7.8	11.0	10.6
Total	$mil	0.0	-4052.9	52.3	71.3	174.0
	%	0.0	43.0	8.2	11.5	21.7
Revenue						
Men	$mil	0.0	288.0	24.1	32.7	83.5
	%	0.0	70.7	12.8	18.2	40.7
Women	$mil	0.0	475.8	53.1	72.5	74.8
	%	0.0	39.4	9.4	13.3	12.7
Total	$mil	0.0	763.9	77.2	105.2	158.2
	%	0.0	47.3	10.3	14.5	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: Robustness: Overstatement of Consumer Welfare Increase

	Percent Increase			**Absolute Increase ($ Millions)**		
Assortment Size	Tailored	National	%Δ	Tailored	National	%Δ
Baseline (b_ℓ)	8.2	11.5	40.2	52.3	71.3	36.3
Threshold						
Mean Baseline ($\bar{b}$)	19.5	28.9	48.7	102.3	138.6	35.5
3000	77.7	109.4	40.9	276.2	330.1	19.5
6000	43.1	62.6	45.1	188.8	239.3	26.7
12000	19.2	28.6	49.0	101.1	137.3	35.8
24000	4.6	7.1	55.4	28.0	41.5	48.1

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 10: Robustness: Overstatement of Retail Revenue

	Percent Increase			**Absolute Increase ($ Millions)**		
Assortment Size	Tailored	National	%Δ	Tailored	National	%Δ
Baseline (b_ℓ)	10.3	14.5	40.8	77.2	105.2	36.3
Threshold						
Mean Baseline ($\bar{b}$)	24.8	37.0	49.4	148.7	199.6	34.3
3000	90.8	127.2	40.0	361.1	425.8	17.9
6000	52.1	75.4	44.7	257.1	321.2	24.9
12000	24.4	36.6	49.6	147.1	197.9	34.5
24000	6.4	10.2	58.7	46.0	68.7	49.4

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.

C An Empirical Bayesian Estimator of Shares

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

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

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

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

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

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

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

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

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

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

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

for an unknown parameter ϑ.

Equations C.1 and C.2 then imply

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

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

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

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

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

Thus,

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

which implies

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

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

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

D Unadjusted Shares and Market-by-Market Analysis

Table 11 reports demand estimates using unadjusted shares at the local and national level (Local NL-US, National NL-US). At the local level, using unadjusted shares appears to result in significant attenuation bias. Of particular concern is the estimated price elasticities of -0.876 and -0.592 for men and women, respectively, are much too small in magnitude. This is driven by the combined attenuation of both the price coefficient and the nesting parameter. However, adjusted shares seems to fair worse. While it does seem to alleviate the attenuation in the nesting parameter, the price coefficient become positive for both men's and women's shoes leading to nonsensical price elasticities. This is likely driven by the sheer number of zeros and the data providing little guidance on how to adjust shares at the local level.

At the national level, where less than 10% of the sample is dropped, we find unadjusted shares yield price elasticities that are very similar in magnitude to estimates using adjusted shares. While the price coefficient is relatively unchanged by adjusting the shares, the nesting parameter is smaller in both then men's and women's specifications when shares are not adjusted. As a result of this attenuation, consumers are estimated have more inelastic demand with unadjusted shares which is consistent with Gandhi, Lu, and Shi (2014).

Another approach of retaining local heterogeneity is to have location-specific parameters, which we operationalize by estimating demand market-by-market. With 32 and 82 million observations for men and women respectively, we found it too computationally intensive to estimate all markets simultaneously. Summary results of these models appear in Table 12. While there is substantial variation in estimates across markets, our general finding is that these models perform poorly with both unadjusted and adjusted shares. For example, the average product level price elasticity using adjusted shares is -1.538 and -0.792 for men's and women's shoes, respectively. Additionally, for adjusted shares, most of the market level price coefficients are positive resulting in nonsensical price elasticities in both specifications.

Table 11: Nested Logit Demand Estimates with Unadjusted Shares

	Men		Women	
	Local	National	Local	National
	(1)	(2)	(3)	(4)
Price	−0.007***	−0.007***	−0.005***	−0.016***
	(0.000)	(0.000)	(0.000)	(0.000)
Comfort	0.033***	0.013***	0.027***	0.050***
	(0.003)	(0.003)	(0.002)	(0.007)
Look	0.000	0.014***	0.006***	0.020**
	(0.003)	(0.004)	(0.002)	(0.008)
Overall	0.045***	0.050***	0.047***	0.159***
	(0.003)	(0.004)	(0.002)	(0.009)
No Review	0.266***	0.296***	0.263***	0.772***
	(0.012)	(0.023)	(0.009)	(0.033)
λ	0.103***	0.777***	−0.034***	0.187***
	(0.010)	(0.009)	(0.003)	(0.011)
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
N	1.8mil	159,280	3.9mil	330,737
Zeroes	95.0%	8.1%	95.0%	8.7%
Price Elast.				
Mean	−0.876	−3.458	−0.592	−2.360
Std. Dev.	(0.619)	(2.446)	(0.474)	(1.891)

Notes: Estimated at the monthly level using empirical (observed) shares. In columns (1) and (3), dependent variables are constructed from local market shares and (2) and (4) dependent variables are constructed from national market shares. Zeros indicate the percentage of products dropped from the sample by using empirical shares. Robust standard errors in parentheses.

Table 12: Summary of Demand Estimates Market-by-Market

	Men		Women	
	Empirical Shares	Adjusted Shares	Empirical Shares	Adjusted Shares
	(1)	(2)	(3)	(4)
Price	−0.002	0.001	−0.001	0.003
	[−0.036, 0.023]	[−0.008, 0.003]	[−0.018, 0.037]	[−0.008, 0.005]
	20.7%**	95.3%**	30.5%**	100.0%**
Comfort	0.008	−0.002	0.008	−0.007
	[−0.389, 0.321]	[−0.006, 0.002]	[−0.233, 0.326]	[−0.013, 0.030]
	13.1%**	63.4%**	18.3%**	94.4%**
Look	0.014	−0.002	0.012	0.003
	[−0.329, 0.773]	[−0.006, 0.017]	[−0.284, 1.276]	[−0.006, 0.012]
	5.2%**	70.9%**	9.4%**	48.8%**
Overall	0.013	0.002	0.017	−0.008
	[−0.357, 0.225]	[−0.004, 0.061]	[−0.341, 0.279]	[−0.021, 0.097]
	12.2%**	50.7%**	20.7%**	93.4%**
No Review	0.124	−0.027	0.125	−0.139
	[−1.154, 1.786]	[−0.095, 0.367]	[−0.943, 3.906]	[−0.232, 0.473]
	19.2%**	93.4%**	22.5%**	99.5%**
λ	0.207	0.982	0.044	0.897
	[−0.366, 1.417]	[0.799, 1.006]	[−0.262, 0.864]	[0.459, 0.992]
	39.0%**	100.0%**	24.4%**	100.0%**
Fixed Effects				
Category	✓	✓	✓	✓
Brand	✓	✓	✓	✓
Color	✓	✓	✓	✓
N	[119, 101,937]	173,304	[391, 210,242]	362,184
Zeros	[41.18, 99.93]	–	[41.95, 99.89]	–
Price Elast.				
Mean	−1.538	1.969	−0.792	1.247
Std. Dev.	(1.935)	(4.558)	(0.873)	(1.288)

Notes: Estimated at the monthly level, market-by-market. Estimate rows are: mean parameter estimates across locations (unweighted), range of estimates, and the percentage of estimates significant at 5%. Columns (1) and (3) use empirical (observed) shares and (2) and (4) use Gandhi, Lu, and Shi (2014) adjusted shares.

E Monte Carlo Analysis

We conduct a Monte Carlo study of our estimator. We start by specifying the data generating process of a nested logit demand system and then create synthetic data sets from this process. Finally, we estimate the structural parameters using 2-step GMM.

The true model specifies consumer utility as

$$\begin{aligned} u_{i\ell j} &= \underbrace{\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \xi_j}_{\delta_j} + \eta_{\ell j} + \zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j} \\ &= -4 + -.75 x_{1j} + .75 x_{2j} + \xi_j + \eta_{\ell j} + \zeta_{ic} + (1-.5)\varepsilon_{i\ell j} \end{aligned}$$

The normalized outside good gives utility $u_{i\ell 0} = \zeta_{i0} + (1-\lambda)\varepsilon_{i\ell 0}$. Here we assume both characteristics are exogenous from the unobservable ξ; however, given real data, instrumental variables can be used on these characteristics. We assign distributions on the data generating process according to Table 13 below.

Table 13: Data generating process for Monte Carlo study

Definition	Variable	Specification
Characteristic 1	x_1	$\mathcal{N}(0,1)$
Characteristic 2	x_2	$\mathcal{N}\left(0, 1.5^2\right)$
National Unobservable	ξ	$\mathcal{N}(0,1)$
Local Unobservable	η	$\mathcal{N}(0, \sigma_c = 1)$
Individual Unobservable	$\zeta + (1-\lambda)\varepsilon$	GEV
Local-Category Product Size	J_c	175
Num. of Categories	C	3
Num. of Periods	T	10
Market Population	M	2000000
Num. of Local Markets	L	200
Population Distribution	ω_ℓ	$1/L$

The parameters to be estimated are: $\beta_0 = -4, \beta_1 = -.75, \beta_2 = .75, \sigma_c = 1, \lambda = .5$. The following steps are used to compute the estimator:

0. Initialize values of σ, λ,
1. Recover $\delta_j^{(k)}$ using the inversion (Equation 2.4),

2. Given, $\delta_j^{(k)}$, calculate GMM objective using micro moments and orthogonality conditions on $\xi_j^{(k)}$, $G(\cdot)$,

3. Select $\sigma^{(k')}, \lambda^{(k')}$ and repeat 1-2 until GMM objective is minimized,

4. Given parameter estimates $\widehat{\theta_1}$, calculate the weighting matrix

$$\widehat{W} = \left(G(\widehat{\theta_1}; Z)^T G(\widehat{\theta_1}; Z)\right)^{-1},$$

5. With $\widehat{W}$ repeat steps 0-3 to obtain $\widehat{\theta_2}$, the two-step feasible GMM estimator.

We minimize to the GMM objective using $Z = [X, z_1, z_2]$ as instruments, where z_k is the mean characteristic of competing products within category for characteristic k. The problem is estimated by calling the solver Knitro using the analytic gradient.

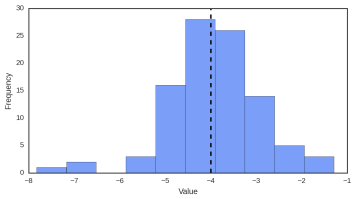
Table 14 presents the results for our Monte Carlo exercises, using 144 synthetic data sets to construct the bias, mean-squared error, and rejection rates. The data generating process yields roughly 75% local zeros and 10% aggregate zeros. We present three sets of Monte Carlo exercises where the micro moments are constructed using: (i) the unconditional share, (ii) the share conditional on purchase, and (iii) the share conditional on category.

Table 14: Monte Carlo Results

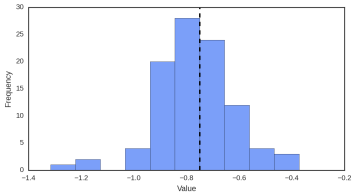
Parameter	True Value	Bias	MSE	Reject. Rates (%)
(i) Unconditional Share				
β_0	-4	0.069	0.018	5.983
β_1	-0.75	0.050	0.049	3.419
β_2	0.75	0.049	0.048	4.274
σ_1	1	0.050	0.049	4.274
σ_2	1	0.043	1.214	4.274
σ_3	1	-0.022	0.028	4.274
λ	0.5	0.022	0.027	4.274
(ii) Conditional on Purchase				
β_0	-4	0.069	0.018	5.983
β_1	-0.75	0.063	0.054	3.419
β_2	0.75	0.062	0.053	3.419
σ_1	1	0.062	0.053	3.419
σ_2	1	0.013	1.231	4.274
σ_3	1	-0.022	0.028	4.274
λ	0.5	0.022	0.027	4.274
(iii) Conditional on Category				
β_0	-4	0.077	0.022	4.274
β_1	-0.75	0.049	0.066	5.128
β_2	0.75	0.044	0.060	4.274
σ_1	1	0.048	0.064	3.419
σ_2	1	0.101	1.477	5.983
σ_3	1	-0.007	0.030	5.983
λ	0.5	0.007	0.030	5.128

The last column tests $H_0 : \widehat{\theta_k} = \theta_0$ and $H_1 : \neg H_0$. The rejection rates are at the 5% level.

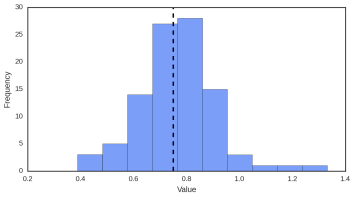
Figure 11: Histogram of Monte Carlo parameter estimates



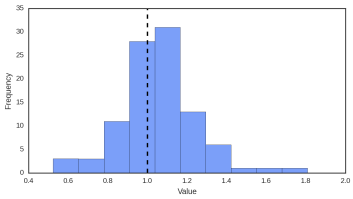
(a) β_0



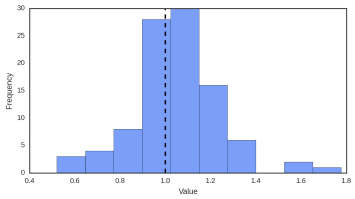
(b) β_1



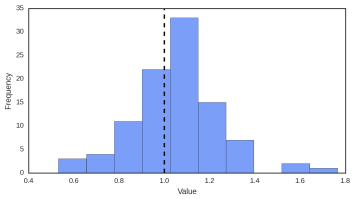
(c) β_2



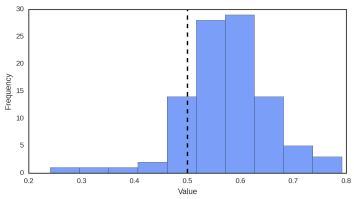
(d) σ_1



(e) σ_2



(f) σ_3



(g) λ