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

Thomas W. Quan  Kevin R. Williams
Department of Economics  School of Management
University of Georgia*  Yale University†

July 2016‡

Abstract

Online retail gives consumers access to astonishing variety of products. However, the additional value created by this variety depends on the extent to which local retailers serve the demands of local consumers. To quantify the gains and account for local demands, we use new, rich data from an online retailer and propose a methodology to address an issue that often arises in such data – sparsity of local sales due to sampling and the significant number of local sales. Our estimates indicate products face substantial demand heterogeneity across geographic markets and, as a result, we find the welfare gains of increasing variety to be between 15% and 40% lower than previous studies.

JEL Classification: C13, L67, L81

*quan@uga.edu
†kevin.williams@yale.edu
‡We thank Steve Berry, Judy Chevalier, Chris Conlon, Amit Gandhi, Phil Haile, Thomas Holmes, Kyoo il Kim, Julie Mortimer, Amil Petrin, Fiona Scott-Morton, Ben Shiller, Catherine Tucker, Joel Waldfogel, and 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 Meetings, 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, and Boston College. We also thank the Minnesota Supercomputing Institute (MSI) for providing computational resources and the online retailer for providing us with data.
1 Introduction

There is widespread recognition that as economies 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 gains from online retail depend critically on the extent to which demand varies across geography and on how traditional brick-and-mortar stores respond to local tastes. 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

\footnote{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.}
this 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 of a commonly purchased good, shoes. We use novel data set from a large online retailer and show abstracting from across-market demand heterogeneity leads to significantly overestimated gains of increasing variety.

Our estimation methodology is necessitated by the characteristics of our data, and of many big 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 are increasingly getting 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 the assumption that the market size is sufficiently large for the observed market share to be observed without sampling error is no longer reasonable.

In practice, these observations 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. Second, the zeros are indicative of a small sample problem. This is particularly problematic for

\(^2\)Aggregated across geographic markets, time, or products.
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 see one shoe sale for a particular market, it would suggest there are no gains to increasing variety because only that 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, 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 overstates the welfare effects because all zeros are adjusted by the same amount (i.e. in Alaska, the unsold boot is adjusted to the same level as the unsold sandal). 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 it 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 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

---

3Another approach is to abandon the GMM framework in favor of maximum likelihood. 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.
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. We flexibly model consumer heterogeneity across markets with random effects and show that a function of our variances corresponds to the Ackerberg and Rysman (2005) penalty term at the aggregate level. 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 which provides a data-driven motivation for its use in applied work.

We use our model to 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 weight 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 variety. In other words, the compensating variation for additional variety is negative and online variety 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 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, a one standard deviation increase in the local demand shock of an average sandal is equivalent to a decline in price of $33. 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 customized 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 between 15% and 40% 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, in our model the random effect 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 $\ell$ 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, $u_{it\ell} \geq u_{itj'}, \forall j' \in J \cup \{0\}$, 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 product every $j$ 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

$$u_{it\ell j} = \delta_{t\ell j} + \zeta_{ic} + (1 - \lambda)\epsilon_{it\ell j}$$

where $\delta_{t\ell j}$ is the mean utility of product $j$ at location $\ell$, $\epsilon_{it\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution and, for consumer $i$, $\zeta_{ic}$ is common to all products in the same category and has a distribution that depends on the nesting parameter $\lambda$, $0 \leq \lambda < 1$. Cardell (1997) shows that $\zeta_{ic} + (1 - \lambda)\epsilon_{it\ell j}$ has a generalized extreme value (GEV) distribution, leading to the frequently used nested logit demand model. The parameter $\lambda$ 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 $\ell$ is linear in product characteristics and can be written as

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

where $x_j$ is a vector of product characteristics, $p_j$ is the price of product $j$, and $\xi_{t\ell j}$ is a
location-specific unobserved product quality. Observable characteristics do not change 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 $\lambda$. 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:

$$
\pi_{\ell,j} = \pi_{\ell,c} \cdot \pi_{\ell,j|c}
$$

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

(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:

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

(2.2)

In the estimation of the standard model, the maintained assumption is that the size of each market $\ell$ 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. In data sets with a large number of products, we may not expect to observe a sale for every product in every disaggregated market. This
suggests the market size assumption may no longer be reasonable. Simple aggregation is also unsatisfactory because it would average over differences across location, obscuring the additional information contained in the disaggregated data. In particular, $\xi_{\ell j}$ will be averaged over the locations aggregated.\footnote{In addition to averaging over $\xi_{\ell j}$, simple aggregation over products would require averaging over the products’ characteristics.}

\section*{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 product mean utilities. We assume that the location-specific unobserved qualities, $\xi_{\ell j}$, are additively separable into two components, an average term that is constant across locations, $\xi_j$, and a location-specific deviation, $\eta_{\ell j}$. Rearranging terms we have,

$$
\delta_{\ell j} = x_j\beta - ap_j + \xi_j + \eta_{\ell j},
$$

where $\delta_j$ 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, $\zeta_{ic} + (1-\lambda)\varepsilon_{ij}$. When $\eta_{\ell j} = 0$ for all $\ell \in L, j \in J$, the model reduces to a 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, \( \omega_\ell \) is the population share of location \( \ell \).

We could invert the market shares for each individual location \( \ell \), 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, the sparsity of individual product sales within locations would lead to a selection bias. To circumvent this problem, we use a random-effects specification, where \( \eta_{\ell j} \) is drawn independently from a normal distribution, \( N(0, \sigma_j^2) \). Instead of attempting to recover each \( \eta_{\ell j} \) directly, we estimate the variance of its distribution, \( \sigma_j^2 \). Like Berry, Levinsohn, and Pakes (1995), our model corresponds to the addition of a random coefficient, where the random coefficient is constant for all consumers within a location. However, unlike BLP, we allow for sampling error in local level shares, which we discuss in the next subsection.

We maintain the fixed effects assumption for \( \xi_j \) which allows for the possibility that prices are correlated with \( \xi_j \). This is useful for applications, such as our own, where prices are set nationally. We assume all product characteristics, including price, are exogenous with respect to \( \eta \).

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,

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

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

(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 \( \eta \) and thus, aggregated choice probabilities only depend on the variance of the across-market heterogeneity. Therefore, national demand can be expressed as

\[
\pi_j = \pi_j(\delta, \lambda; \sigma), \ j = 1, \ldots, J,
\]

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\(^5\)

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

Equation 2.4 relates \( \delta_j \) to the aggregated share data, \( \pi_j \), local population shares, \( \omega_\ell \), 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 \( \eta_{\ell j} = 0, \)

\(^5\)See Appendix A.
$\forall \ell \in L, j \in J$. 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

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

where, for each location $\ell$, the expectation is over all products $j$. The complexity of this expectation is highlighted when we apply the Law of Iterated Expectations,

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

The conditional expectation not only depends on the 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 make further progress. We approximate the conditional expectation since $\mathbb{E} \left[ \exp \left\{ \frac{\eta_{\ell j}}{1-\lambda} \right\} \bigg| \pi_{\ell c}, \pi_{0\ell} \right] \to \mathbb{E} \left[ \exp \left\{ \frac{\eta_{\ell j}}{1-\lambda} \right\} \right]$ as the number of products tends toward infinity. The unconditional expectation is simple to compute using the moment generating function of the normal

---

$^6$Suppose $\eta_{\ell j} = 0$, $\forall \ell \in L, j \in J$, then $\pi_{0\ell} = \pi_0$ and $\pi_{c\ell} = \pi_c$, and

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

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

$$= (1-\lambda) \log \pi_j + \lambda \log \pi_c - \log \pi_0$$

$$= \log \pi_j - \log \pi_0 - \lambda \log \left( \frac{\pi_j}{\pi_c} \right)$$

$$= \log \pi_j - \log \pi_0 - \lambda \log \pi_{jk}$$
distribution,\(^7\) \(E\left[ \exp \left\{ \frac{\eta_{j\ell}}{1-\lambda} \right\} \right] = \exp \left\{ \frac{1}{2} \frac{\sigma_j^2}{(1-\lambda)^2} \right\} \). This term does not depend on the location. 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 will average out. We demonstrate with Monte Carlo exercises that using the unconditional expectation to approximate the conditional expectation performs well (see Appendix D).

Finally, while small sample sizes make observed local market shares unreasonable estimates of the true underlying choice probabilities for individual products, we assume the local choice probabilities of the outside good, \(\pi_{\ell0}\), and the category shares, \(\pi_{\ell c}\), are well estimated by the data. This is reasonable if the number of categories is small relative to the number of products, and the size of the population of large relative to the number of categories. With this assumption and given \((\sigma, \lambda)\), we can then recover national mean utilities as function of observables \((\pi_j, \pi_{\ell c}, \pi_{\ell0})\),

\[
\delta_j = (1 - \lambda) \left( \log(\pi_j) - \frac{1}{2} \frac{\sigma_j^2}{(1-\lambda)^2} - \log \left( \sum_{\ell \in L} \omega_{\ell} \pi_{\ell c} \left( \frac{\pi_{\ell0}}{\pi_{\ell c}} \right)^{\frac{1}{1-\lambda}} \right) \right). \tag{2.5}
\]

The large number of products has two important implications. First, \(\delta_j\) can be recovered point-wise rather than requiring simultaneously solving a \(J \times J\) system of equations for each location \(\ell\). This greatly reduced the computational burden of the problem. Second, at the national level, it suggests an adjustment that corresponds to the crowding penalty term proposed in Ackerberg and Rysman (2005). Define

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

\(^7\)The moment generating function of a normal distribution with mean \(\mu\) and variance \(\sigma^2\) is,

\[
E \left[ \exp \left\{ \mu t \right\} \right] = M_r(t) = \exp \left\{ \mu t + \frac{1}{2} \sigma^2 t^2 \right\}.
\]
Since $R(\sigma_j)$ is not indexed by $\ell$, the share equation can be rearranged to yield

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

Expanding this equation, we obtain

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

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). To complete the link, it must be that $R(\sigma_j) \to \infty$ as $J \to \infty$.

Holding fixed $\pi_j$ and $\delta_j$, consider the effect on $R(\sigma_j)$ as the set of products $J$ increases. From Equation 2.6, we can see that the denominator grows unboundedly as additional products are added, which would lead to $\pi_j$ to converge to 0. In order to hold $\pi_j$ fixed it must be that $\sigma_j \to \infty$.

Therefore, our random effects corresponds to the Ackerberg and Rysman (2005) crowding penalty term at the aggregate level. While they operationalize the crowding penalty term by making ad-hoc assumptions on the number of retail outlets per product, we are able to motivate and discipline the penalty by examining across-market demand heterogeneity. We estimate the crowding terms using micro-data, as shown in the next subsection.

The incorporation of random effects has important implications for our demand esti-

---

\( ^8 \)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

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

Finally, substituting back in for $\tilde{\delta}_j = \delta_j + (1 - \lambda) \log R(\sigma_j)$ gives us Equation 2.6.
mates. 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 welfare impact of each subsequent product entry. In our model, at the underlying local market level, a large \( \sigma_j \) suggests that demand for that product is highly concentrated in particular geographic markets. The larger \( \sigma_j \) 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 \( \sigma_j \)s are higher because fewer consumers are affected.

2.4 Micro Moments

To identify the random effects, we need additional moments that capture the differing degrees of across-market heterogeneity among products. Zero sales are viewed as problematic when estimating demand. Here, we appeal to them as the source of identification. Let \( P_0_{\ell j}(\sigma; \delta, \lambda) \) be the probability that a product \( j \) has zero sales, given \( N_\ell \) consumers are observed to make any purchase at location \( \ell \). We then define

\[
P_0_j(\sigma; \delta, \lambda) = \frac{1}{L} \sum_{\ell=1}^{L} P_0_{\ell j}(\sigma; \delta, \lambda)
\]

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 \( \delta \) as \( \delta(\pi, \lambda, \sigma) \). The empirical analogue is

\[
\hat{P}_0_j = \frac{1}{L} \sum_{\ell=1}^{L} 1\{s_{\ell j} = 0\},
\]

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

\[
mm_{j}(\sigma; \delta, \lambda) = \left( P_0_j(\sigma; \delta, \lambda) - \hat{P}_0_j \right).
\]
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.\(^9\)

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

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

where $\bar{\eta}_{\ell j}$ is an i.i.d. draw from a standard normal distribution. With the assumptions on the individual level unobservable (GEV), there is a closed form for the local-level product choice probabilities ($\hat{\pi}_{\ell j}$) for any candidate of $\sigma$ and $\lambda$. The local level choice probabilities can be used to simulate consumer purchases at each location, holding the number of observed purchases, $N_{\ell}$, fixed. In particular, the probability a product is observed to have zero sales at location $\ell$ is

$$P_{0\ell j}(\sigma; \delta, \lambda) = (1 - \hat{\pi}_{\ell j})^{N_{\ell}}$$

\(^9\)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. We estimate these models as well.
i.e. the probability we observe $N_\ell$ sales, none of which are good $j$. We then average over locations and match it to the fraction of locations observing zero sales of $j$. We will denote these micro moments, $mm(\cdot)$. This approach is computationally fast and avoids the problems posed by simulating individual purchase decisions.

With a candidate solution of $\sigma$ and $\lambda$, the structure we have placed on the $\eta$s 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 $\delta$ as defined in Equation 2.5,

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

because $\xi_j$ enters the aggregate product share, $\pi_j$, and the local level category shares, $\pi_\ell c$. Berry (1994) solves this problem by employing an instrument, $z_{j|c}$, that is correlated with the within category share, but uncorrelated with the unobserved product quality. 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.

---

10Alternatively, another way to formulate the micro moment, taking local category shares as given, is to match the probability of zero sales within category, $P0_{\ell|c}(\sigma; \delta, \lambda) = (1 - \hat{\pi}_{j|c})^{N_\ell c}$, where $N_\ell c$ is the number of purchases within category $c$ at location $\ell$.

11For example, a combination of the product characteristics of competing products within the same category or nest.
Let \( Z \) be the usual matrix of nested logit instruments that identify \( \beta, \alpha, \lambda \) and denote the set of moments, \( m = E[Z'\xi] \). Stacking our moments and micro moments where \( \theta = (\sigma, \lambda, \beta, \alpha) \) we have
\[
G(\theta; \cdot) = \begin{bmatrix} \text{mm} \\ m \end{bmatrix}
\]
and the GMM criterion is \( G(\theta; \cdot)^TWG(\theta; \cdot) \), with weighting matrix \( W \). We first take \( W = I \) and then use \( \hat{W} = (G(\hat{\theta}(1); \cdot)G(\hat{\theta}(1); \cdot)^T)^{-1} \) in the second step. Our final estimates are \( \hat{\theta}(2) \).

### 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 localization of assortments using the brick-and-mortar assortment data (Section 3.2) and then demand-side across-market heterogeneity using the online 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. 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

---

12Initially 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.
data is size specific, the sales data does not include size. We concede that this, in general, will cause us to understate the gains of online variety because consumers with unusual foot sizes may greatly benefit from online shopping if traditional retailers do not typically stock unusual sizes. 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 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 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. The only information we obtain on the shoe is the identifier number. We do not observe for example the shoe brand or type. 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; however, we can use the data to examine local assortments.

---

13 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. Using our Macy’s data, we reject the null hypothesis that the mean assortment shoe size is constant across stores.

14 We did not scrape webpages but rather downloaded targeted information via the sites’ APIs.
Table 1 presents summary information on the assortments of 649 Macy’s locations and 3,141 Payless’ locations. In September 2014, we observe 7,844 different styles available at Macys.com, of which about 35% 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 similar across retail chains – 624.9 and 513.0 for Macy’s and Payless, respectively. However, there is greater variance in Macy’s store size. 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

The premise of this paper is that there may exist significant differences in consumer demand across geographic markets. This has significant implications for the value of increased product variety if traditional brick-and-mortar retailers localize assortments. For large national retailers, there are trade offs to localizing assortments. On the one hand, catering to local demand may greatly increase revenues, but on the other hand, there are cost advantages from economies of scale through standardization. Available evidence suggests the former may outweigh the latter. For example, in recent years, Macy’s has made a concerted effort to better localize its product assortments through a program called "My Macy’s:"

"We continued to refine and improve the My Macy’s process for localizing merchandise assortments by store location.... We have re-doubled the emphasis on precision in merchandise size, fit, fabric weight, style and color preferences by store, market and climate zone. In addition, we are better understanding and serving the specific needs of multicultural consumers who represent an increasingly large proportion of our customers."\(^1\)

\(^1\)Ghemawat (1986) found 70% of Walmart’s merchandise was common across stores, and 30% was tailored to local needs.

\(^\text{16}\) https://www.macysinc.com/macys/m.o.m.-strategies/default.aspx
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 with similar sizes. 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.

Overall, we interpret the general trends presented above as of firms responding to demand, but we acknowledge that there are likely 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,\textsuperscript{17} in addition to some locations where retailers created 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.

\textsuperscript{17}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.
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). 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 market revenue share and the national revenue share. This suggests that the commonality of popular products is quite small across markets.

We formally test for across-market demand heterogeneity, controlling for local sample size, using multinomial tests comparing local market shares \( s_\ell \) to national market shares \( s \). 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

---

18 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.

19 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.
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 exploit this variation to help us identify across-market demand heterogeneity.

3.4 Aggregation and the Zeros Problem

If we define local market shares given observed local sales, the vast majority of products would have 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). Our market sizes are defined as 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, top brands, and time. We instrument for both price and the nesting parameter using the typical BLP-style instruments. Included are the number of available styles (color combinations) for a particular shoe model, and average characteristics, excluding that product, within a brand. Since the review covariates are highly collinear, we use only the average overall rating within a brand-category as instruments. 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

\[
\sum_{j' \neq j} x_{j'}, \quad \sum_{j' \neq j \in c_b} x_{j'}.
\]

These will aid us in identifying the price coefficient, \( \alpha \), and the nesting parameter, \( \lambda \).

In principle, with our modeling assumptions and a large number of product-location

\[\text{20}^{th}\text{We find at finer levels of geography, such as zip code, the nearly 100\% local zeros cause the micro moments to lose identifying power. We have confirmed this with Monte Carlo exercises, some of which appear in the Appendix. We choose CSA+state, compared to just CSA, since a large percentage of observed sales occur outside CSAs. For example, if we pursued the CSA market definition, we would drop all of sales to consumers in Alaska. Results dropping states are available on request.}\]
observations, we could estimate a $\sigma_j$ 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 $\sigma$ as

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

meaning $mm()$ contains $C$ moments.\footnote{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.}

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

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

We estimate the Local FE model for two sets of data. The first treats observed shares as true shares and drops all of the zero observations. 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 FE. Results for US shares are considered in the Robustness section that follows.

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. A detailed discussion of the correction procedure can be found in Appendix C.
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 three sets of estimates, corresponding to: (1) Local FE - US; (2) Local FE-AS; (3) National FE-AS; and (4) Local RE. For each of our specifications, we also compute individual product level price elasticities. For the national level estimates, price elasticities are computed as

\[ e_j = \frac{\partial \log \pi_j}{\partial \log p_j} = \alpha p_j \left( \frac{1}{1 - \lambda} - \frac{\lambda}{1 - \lambda} \pi_{jk} - \pi_j \right). \]

For the local level estimates, these elasticities are computed first at the location-product level. Then the weighted average, by market size, of the location-product level elasticities are taken to obtain an aggregate individual product level elasticity. That is

\[ e_j = \sum_{\ell} \omega_{\ell} \frac{\partial \log \pi_{\ell j}}{\partial \log p_{\ell j}}. \]

Specifications (1) and (2), Local FE-US and Local FE-AS, illustrate the selection bias generated by the severity of the zeros problem, even when employing adjusted shares in the latter case. In these models, each observation is a product-location specific share; thus, the number of observations is 213 times greater (number of products times 213 CSA+state markets) than the other two specifications. Unfortunately, at this level of disaggregation, about 95% of the observations have zero sales. 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. The impact of attenuation bias in the price or the nesting parameter imply price elasticities that are much too inelastic. The bottom panels of each table report the mean and standard deviation of the estimated product level price elasticities. We obtain elasticities of -0.9 and -0.8 for men using US and AS data, respectively, which
differs by a magnitude of about four compared to the other models. We also fine a large bias toward zero for women. The distributions for the estimated price elasticities are quite skewed with fat left tails (more elastic), leading to large standard deviations. However, since \( \alpha < 0 \) and \( \lambda \in (0, 1) \), none of our products are estimated to have upward sloping demand curves.

Specifications (3) and (4) directly compare the results estimated using the standard approach on national level data and the results estimated using our procedure for the nested logit model. The estimates of the price coefficient and nesting parameter are similar and significant, which yields average product level price elasticities of \((-3.6, -3.0)\) and \((-2.1, -1.5)\) for men and women, respectively, across specifications.\(^{22}\) The reasonably high estimates of the nesting parameter suggest substitution within category is important. The similarity of the results for these two specifications is unsurprising – they purportedly are estimates of the same aggregate mean utility. However, the advantage of utilizing the Local RE model is that it retains information on the distribution of heterogeneity across locations. The importance of this distinction will be highlighted in the following section when we perform counterfactual analyses.

Turning to the coefficients on our review variables, we can see that the overall rating has the expected sign, with higher ratings having positive effects on demand. The estimates are statistically significant. Look and comfort have much smaller effects, with only comfort being statistically significant in the women’s specification. The review ratings are highly correlated so it is likely that after controlling for overall appeal, the estimates for look and comfort suffer from collinearity. 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

\(^{22}\)The empirical literature on shoe demand is limited. Roberts, Xu, Fan, and Zhang (2012) look at imports of Chinese footwear and find elasticities in the range that we find. For the US, their elasticities are smaller; however, their definition of a shoe is broader than our study.
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

The results in the previous subsection suggest the Local RE model provides comparable estimates to the National FE-AS model. However, the Local RE model also provides estimates of across-market demand heterogeneity to rationalize the distribution of local demand. Our estimates for the across-market demand heterogeneity in the Local RE model are presented in Table 7.

We find roughly half of the across-market heterogeneity parameters to be statistically significant. The statistically significant parameters are highly significant economically. For example, the smallest statistically significant $\sigma_c$ for men’s shoes is sandals at 0.46. To put this number in perspective, a one standard deviation increase in a sandal’s draw of $\eta_{\ell j}$ is equivalent to a decline in price of $33. Similarly, for women’s shoes, loafers have the smallest statistically significant $\sigma_c$ at 0.32, which implies a one standard deviation increase in a loafer’s draw of $\eta_{\ell j}$ is equivalent to a decline in price of almost $46$. These large effects will have important implications for consumer welfare analysis, as we will see in the following section.

To show how the estimates of $\sigma_c$ rationalize the distribution of local sales in the data, Figure 4 plots the proportion of local zero market shares across products in the sneaker category for three scenarios: (i) the observed data, (ii) results of the Local RE model, and (iii) results of the Local RE model for the case of homogeneous demand across markets, i.e. when $\sigma_j = 0$ for all shoes $j \in J$. The left panel is the plot for men’s sneakers and the right panel is for women’s sneakers. For women’s sneakers in particular, we see the Local RE model closely follows the data, which makes sense since the micro moments are used to match local zeros. However, we see the homogeneous demand line systematically understates the degree of zeros. Put another way, the gap between the data line and
homogeneous demand line suggests that if demand were homogeneous across markets, we would expect to see far fewer zeros among popular and mid-ranked products.

Figure 4 also suggests that there is a limit to the level of disaggregation the Local RE model can handle. In particular, it becomes increasingly difficult to identify the variance in demand when markets have too few observed sales because any pattern of zeros can be rationalized through the mean aggregate utilities. For example, in the extreme, we could define our local markets as individual consumers. However, we would expect to estimate no “across-market heterogeneity” because it would already be captured by the mean utilities and the logit error.

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 our estimated 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 $\eta$'s 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, location 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 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 level assortment sizes. Each market receives a limited assortment from the predicted values of

\[
\log(a_\ell) = \beta_0 + \beta_1 \log(\text{pop}_\ell) + \epsilon_\ell,
\]

where \(a_\ell\) is the assortment number we see in our Macy’s and Payless data and \(\text{pop}_\ell\) is location population. We examine the robustness of our results for a range of thresholds in the following section. Location level consumer surplus is defined 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

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

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

Table 10 summarizes our main findings and compares estimates under the Local RE model with the Local FE-ES model. Using our proposed technique, we find consumer welfare increases by $16.8 million, or 9.9%, for men and $106 million, or 15.4%, for women by going from localized assortments to the large, online choice set. Our interpretation is that these numbers are sizable, but, as we will see, they are significantly smaller than gains
under a nationally standardized assortment.

These numbers are also in stark contrast to the gains found under the local fixed effects model. Consistent with Ellison and Glaeser (1997), 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 most consumer demand. This leads to an understatement of the consumer welfare increase due to having access to the large online choice set, and in fact, we estimate the gains to be almost nonexistent. For men’s shoes, we find the consumer gain to be approximately 0% and for women, 0.1%.

To get a sense of the role that across-market demand heterogeneity and localized assortments play in the counterfactual, we now compare these results to the scenario where each location draws from the same, standardized ranking of products. Each location receives a number of products based on its population, the same number as in the previous counterfactuals, but the selection of products is derived from the top products at the aggregate, or national, level. Under a standardized national ranking of products, access to online variety increases consumer welfare by $23.3 million, or 14.3%, for men and $121.7 million, or 18.1%, for women. This suggests abstracting from the fact that retailers cater to local demand has sizable consequences concerning the benefit of variety. In absolute terms the overstatement is 38.7% and 14.8% for men and women, respectively. Similarly, in percentage terms the overstatement is 44.4% and 17.5% for men and women, respectively. 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 rates sneakers and sandals, there will be too few boots for consumers in Alaska.

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 3% higher under the latter. This suggests that there
may be an incentive for local stores to cater to local demand depending on the potential diseconomies of scale due to localization.

5.2 Long Tail Analysis

Our counterfactual results suggest 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 median market (by number of monthly sales), middle 10% (p45-p55), middle 50% (p25-p75), and national level markets. 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. Using the middle 50% of markets (p25-p75), the lengthening effect of the tail is very pronounced. Hence, simply aggregating over markets creates a long tail even though each individual market demands far fewer varieties of shoes.

On the other hand, the small sample problem in the raw data presents us with a skewed perspective in that it suggests a ridiculously short tail at the local level. However, we can correct for the small sample problem in our long tail analysis by utilizing the results of our 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. As expected, we find that the local tail is actually quite a bit longer than suggested by the raw data. This implies that there is a long tail
effect as described in the existing literature, but it is shorter than the national tail would suggest.

Table 11 further illustrates the effect of small sample sizes on the local tail. It presents the average share of revenue accruing to products outside of the top 3,000 products at both the local and national level. At the national level, more than 50% of revenue comes from products ranked outside the top 3,000. At the local level, if we were to rely on the raw data, we would find that only 3.6% of revenue comes from products ranked outside the top 3,000. In other words, 96.4% of demand could be satisfied with just 3,000 well-targeted products in each local market. Simulating our model with the same small number of sales observed in each local market yields a distribution of sales that is similar to the data. However, by simulating a large number of sales in each local market, we find that there is, in fact, significant demand for niche products at the local level. On average, 47.0% of sales come from products ranked outside the top 3,000 products. Consistent with Figure 6, however, the local percentage is less than in the national level data.

6 Robustness

6.1 Welfare and Counterfactual Choice Sets

In this subsection, we examine the robustness of our findings to the size of the counterfactual choice set. While we find that the absolute size of the overstatement is sensitive to the size of the counterfactual assortment size, the percentage overstatement is fairly robust across a wide range of threshold sizes and in line with our findings from the previous section. Table 12 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.

23Given the small number of sales at the local level, this result is unsurprising. For example, suppose fewer than 3,000 sales are observed in a local market. Then, of course, the share of revenue going to products outside the top 3,000 is zero.
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 7, 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 13 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 5.1% 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 8 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.

6.2 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 9 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 profitability. To perform this analysis, we develop new methodology that allows us to confront the severe small sample problem in our data, while retaining 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 one hand, differences in local demand may create an incentive for retailers to customize assortments and our brick-and-mortar data suggests that brick-and-mortar shoe stores are reacting to these incentives. Our results suggest local retailers may generate 3% additional revenue by localizing assortments. On the other hand, our calculations suggest that abstracting from across-market demand heterogeneity overestimates the consumer welfare gain due to online markets by about 20%. This is important since there are several potential avenues through which online retail can benefit consumers. For example, the entry of online firms suggests local brick-and-mortar retailers now face increased competition, which may 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 this channel may be less important than previously thought due to correlated local preferences.

Finally, our results suggest a new interpretation of the long tail phenomenon. With our data, we show that simply aggregating local demand creates a long tail at the national level. We confirm that substantial heterogeneity exists at the national level, but that the key driver is the aggregation of across-market demand heterogeneity.

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 do not observe consumer search which may further decrease the consumer surplus gains of the large online choice set. However, recommendation and search tools available on most online retail websites decrease these search costs and may provide additional benefit compared to physically searching (shopping) at brick-and-mortar stores. Incorporating this feature of the market may be important given the size of the choice set.

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. It may be that the internet has shaped how retailers choose assortments, which in turn has impacted how consumers differentially shop online versus at stores. Additionally, we assume that brick-and-mortar retailers are able to perfectly predict consumer demand. To the extent that buyers for the brick-and-mortar retailers are unable to perfectly forecast local demand, our counterfactuals will understate the gains from online variety.\footnote{Aguiar and Waldfogel (2015) explore the benefit of having a long tail of products when quality is not perfectly predictable.} However, as long as there was some degree of local customization in brick-and-mortar assortments before the Internet and as long as brick-and-mortar retailers are not stocking products at random today, our main conclusion holds. That is, it is important to account for across-market demand heterogeneity when
estimating the gain from online variety.

On the methodological side, our locations are simply characterized by random draws from the distribution of across-market heterogeneity. This means we cannot conduct location-specific analyses. A potentially interesting area of future research is addressing sampling in more flexible demand systems incorporating location-level or individual-level data.

References


A Proofs

Proposition 1. For each product \( j \in J \), applying the law of large numbers in \( L \) and integrating out over \( \eta \) gives

\[
\sum_{\ell \in L} \omega_{\ell} \pi_j (\eta_{\ell}; \delta, \lambda) - \pi_j \rightarrow_{a.s.} 0 \tag{A.1}
\]

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

\[
\pi_j (\eta_{\ell}; \delta, \lambda) = \frac{\Pr(\ell \in c) \cdot \Pr(j | c)}{\sum_{c} \Pr(\ell | c)}.
\]

Aggregating over local choice probabilities gives

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

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 \( \ell \) is equal to

\[
\pi_{\ell 0} = \frac{1}{1 + \sum_{c \in C} D_{\ell c}^{\delta_j + \eta_{\ell j}}},
\]

and note that the probability of choosing a good in category \( c \) at location \( \ell \) is equal to

\[
\pi_{\ell c} = \frac{D_{\ell c}^{\delta_j + \eta_{\ell j}}}{1 + \sum_{c \in C} D_{\ell c}^{\delta_j + \eta_{\ell j}}},
\]

thus

\[
D_{\ell c} = \left( \frac{\pi_{\ell c}}{\pi_{\ell 0}} \right)^{1/\delta_j + \eta_{\ell j}}.
\]

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

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

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,

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

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

$$\sum_{i \in \mathcal{C}} \omega_i \mathbb{E} \left[ \pi_j(t; \delta, \lambda) \right] = \mathbb{E} \left[ \pi_j(t; \delta, \lambda) \right] = \pi_j$$

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 $\left( \frac{1}{\pi \ell_j} \right)^K$ has finite mean and variance, for all $K \geq 0$.\(^{25}\)
- It is then easy to show that $\left( \omega_i \pi \ell_i \left( \frac{\eta_i}{\eta} \right)^m \exp \left\{ \frac{\eta_i}{\eta} \right\} \right)$ 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,

$$\left( \frac{1}{\pi \ell_j} \right)^K = \left( \frac{\sum c \, D_{\ell c}^{-\lambda}}{D_{\ell c}^{1-\lambda}} \right)^K$$

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

$$\left( \frac{\sum c \, D_{\ell c}^{-\lambda}}{D_{\ell c}^{1-\lambda}} \right)^K \leq \left( \left| \mathcal{C} \right| \frac{\sum c \, D_{\ell c}^{-\lambda}}{D_{\ell c}^{1-\lambda}} \right)^K$$

where $|\mathcal{C}|$ denotes the cardinality of $\mathcal{C}$, i.e. the number of categories. This step is achieved by applying Jensen's Inequality of the form $|a_i|^{\ell} : a_i > 0, \forall i$

$$(a_1 + \ldots + a_k)^k \leq k^{k-1}(a_1^k + \ldots + a_k^k)$$ if and only if $k \geq 1$,

where the inequality is reversed if $k \leq 1$. Let $m = (|\mathcal{C}|)^K$, we have

$$\left( \left| \mathcal{C} \right| \frac{\sum c \, D_{\ell c}^{-\lambda}}{D_{\ell c}^{1-\lambda}} \right)^K = m \left( \left( \sum c \, D_{\ell c}^{-\lambda} \right)^{1-k} \right)^K = m \left( \sum c \, D_{\ell c}^{-\lambda} \right)^K,$$

where we let $K = (1 - \lambda)$. The numerator is equal to $\sum_{c, \ell} \exp \left\{ \frac{\eta_i + \eta_j}{\eta} \right\}$, i.e. summed over all products, and the denominator is similarly defined but summed over only $j \in c$. Define $\tilde{N}_i = \frac{\eta_i + \eta_j}{\eta}$ and let good 1 be an arbitrary good in category $c$. Then we have

$$m \left( \sum c \, D_{\ell c}^{-\lambda} \right)^K = m \left( \sum_{c, \ell} \exp \left( \tilde{N}_i \right) \right)^K \leq m \left( \frac{\sum_{c, \ell} \exp \left( \tilde{N}_i \right)}{\exp \left( \tilde{N}_1 \right)} \right)^K.$$

Suppose $K \leq 1$. Then by Jensen’s Inequality,

$$m \mathbb{E} \left[ \sum_{c, \ell} \exp \left( \tilde{N}_i \right) \right] \leq m \left( \mathbb{E} \left[ \sum_{c, \ell} \frac{\exp \left( \tilde{N}_i \right)}{\exp \left( \tilde{N}_1 \right)} \right] \right)^K.$$

\(^{25}\)This proof relies on $f$ being finite; however, that assumption is not necessary for the result to hold. Instead, applying expectations and the moment generating function for the normal distribution also establishes the result.
We have
\[
E \left[ \sum_{j} \frac{\exp(N_j)}{\exp(N_i)} \right] = E \left[ \sum_{j} \exp(N_j - N_i) \right] = \sum_{j} E \left[ \exp(N_j - N_i) \right].
\]
This is a finite sum of the expectations of random variables, each with finite mean. Hence,
\[
E \left[ \left( \frac{1}{\pi \ell_c} \right)^k \right] \leq m \left( \sum_{j} E \left[ \exp(N_j - N_i) \right] \right)^k < \infty,
\]
and raising this sum to the \( \eta \)th power is also finite.

Suppose \( K > 1 \). By Jensen’s Inequality, we have
\[
E \left[ \sum_{j} \frac{\exp(N_j)}{\exp(N_i)} \right] ^{\frac{1}{\ell_c}} \leq \left( \sum_{j} E \left[ \exp(N_j - N_i) \right] \right)^{\frac{1}{\ell_c}}.
\]
The right hand side is equal to
\[
j^{K-1} \sum_{j} E \left[ \exp(KN_j - K\bar{N}_1) \right],
\]
which, again, is a finite sum of expectations of random variables with log-normal distributions, each with finite mean. Hence,
\[
E \left[ \left( \frac{1}{\pi \ell_c} \right)^k \right] \leq j^{K-1} \sum_{j} E \left[ \exp(K\bar{N}_j - K\bar{N}_1) \right] < \infty
\]
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 \( \omega \eta \pi \ell_c \left( \frac{\pi_0}{\pi_0} \right)^{\frac{1}{\ell_c}} \) is a random variable bounded between 0 and 1 and \( \exp \left( \frac{\eta_i}{1-\lambda} \right) \) is a lognormal random variable, it is the straightforward to show that \( \left( \omega \eta \pi \ell_c \left( \frac{\pi_0}{\pi_0} \right)^{\frac{1}{\ell_c}} \exp \left( \frac{\eta_i}{1-\lambda} \right) \right) \) has finite mean and variance. Note also that their expectations only differ in \( \omega \eta \), which is bounded between 0 and 1. Thus, the variance is bounded above when \( \omega \eta = 1 \) and
\[
\sum_{l=1}^{\infty} \frac{V[A \eta \exp \left( \frac{\eta_i}{1-\lambda} \right)]}{\ell_c^2} < \sum_{l=1}^{\infty} \frac{V[A \eta \exp \left( \frac{\eta_i}{1-\lambda} \right)]}{\ell_c^2} < \infty.
\]
Therefore, by Kolmogorov’s Strong Law of Large Numbers
\[
\exp \left( \frac{\delta_i}{1 - \lambda} \right) \sum_{l=1}^{\infty} \omega \eta \pi \ell_c \left( \frac{\pi_0}{\pi_0} \right)^{\frac{1}{\ell_c}} \exp \left( \frac{\eta_i}{1-\lambda} \right) - \exp \left( \frac{\delta_i}{1 - \lambda} \right) \sum_{l=1}^{\infty} \omega \eta \pi \ell_c \left( \frac{\pi_0}{\pi_0} \right)^{\frac{1}{\ell_c}} \exp \left( \frac{\eta_i}{1-\lambda} \right) \rightarrow_{a.s.} 0.
\]
Since \( \delta_i \) is a constant, rearranging it back into the expectation obtains our result. ■
B Tables and Figures

Figure 1: Assortment Overlap by Distance

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{figure1}
  \caption{Assortment Overlap by Distance}
  \label{fig:assortment_overlap}
\end{figure}

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: Goodness of Fit: Percentage of Location Level Zeros (Sneakers)

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

Figure 5: Aggregating to the Long Tail

Note: For varying levels of aggregation, the cumulative share of revenue going to the top products.
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).
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 8: 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 9: Demand Aggregation Dropping Single Sale Observations.

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

<table>
<thead>
<tr>
<th></th>
<th>Macy’s</th>
<th>Payless Shoes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stores</td>
<td>649</td>
<td>3,141</td>
</tr>
<tr>
<td>Number of products</td>
<td>7,844</td>
<td>1,430</td>
</tr>
<tr>
<td>Percent online exclusive</td>
<td>34.8%</td>
<td>19.2%</td>
</tr>
<tr>
<td>Avg. assortment size</td>
<td>624.9</td>
<td>513.0</td>
</tr>
<tr>
<td></td>
<td>(299.3)</td>
<td>(58.4)</td>
</tr>
</tbody>
</table>

Notes: Data collected through macys.com and payless.com. For every shoe-size combination, we check to see if the product is in stock. \( N_{\text{Macy’s}} = 93,481,515 \), \( N_{\text{Payless}} = 69,451,866 \).

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

<table>
<thead>
<tr>
<th>Market Definition</th>
<th>Number of Markets</th>
<th>Market Top 500</th>
<th>Multinomial Tests - Rejection Rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-Digit Zip Code</td>
<td>35,279</td>
<td>99.96</td>
<td>4.69</td>
</tr>
<tr>
<td>3-Digit Zip Code</td>
<td>894</td>
<td>85.12</td>
<td>6.28</td>
</tr>
<tr>
<td>CSA + State</td>
<td>213</td>
<td>67.05</td>
<td>7.23</td>
</tr>
<tr>
<td>Combined Statistical Area (CSA)</td>
<td>165</td>
<td>70.31</td>
<td>7.19</td>
</tr>
<tr>
<td>State (plus DC)</td>
<td>51</td>
<td>30.04</td>
<td>9.86</td>
</tr>
<tr>
<td>Census Region</td>
<td>4</td>
<td>16.36</td>
<td>14.76</td>
</tr>
<tr>
<td>National</td>
<td>1</td>
<td>15.54</td>
<td>15.54</td>
</tr>
</tbody>
</table>

Multinomial tests: Define \( s = [s_1, ..., s_J] \) and \( s_t = [s_{t1}, ..., s_{tJ}] \), then the null hypothesis is \( H_0 : s = s_t \). 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

<table>
<thead>
<tr>
<th>Market Definition</th>
<th>Number of Markets</th>
<th>Percent of Zero Sales</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Week</td>
<td>Month</td>
<td>Quarter</td>
<td>Annual</td>
<td></td>
</tr>
<tr>
<td>3-Digit Zip Code</td>
<td>894</td>
<td>99.57</td>
<td>98.57</td>
<td>97.07</td>
<td>94.09</td>
<td></td>
</tr>
<tr>
<td>CSA + State</td>
<td>213</td>
<td>98.43</td>
<td>95.54</td>
<td>91.98</td>
<td>86.12</td>
<td></td>
</tr>
<tr>
<td>Combined Statistical Area (CSA)</td>
<td>165</td>
<td>98.50</td>
<td>95.80</td>
<td>92.53</td>
<td>87.15</td>
<td></td>
</tr>
<tr>
<td>State (plus DC)</td>
<td>51</td>
<td>94.23</td>
<td>85.25</td>
<td>76.27</td>
<td>64.26</td>
<td></td>
</tr>
<tr>
<td>Census Region</td>
<td>4</td>
<td>59.83</td>
<td>33.70</td>
<td>21.72</td>
<td>12.17</td>
<td></td>
</tr>
<tr>
<td>National</td>
<td>1</td>
<td>28.30</td>
<td>9.27</td>
<td>4.50</td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>

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

Table 4: Revenue Share of Top Products with Product Aggregation

<table>
<thead>
<tr>
<th>Product Definition</th>
<th>Percent of Zero Sales</th>
<th>Market Top 500 Market</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKU (shoe + style)</td>
<td>95.54</td>
<td>67.05</td>
<td>7.23</td>
</tr>
<tr>
<td>Shoe</td>
<td>93.10</td>
<td>73.39</td>
<td>19.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Market Top 10 Market</td>
<td>National</td>
</tr>
<tr>
<td>SKU (shoe + style)</td>
<td>95.54</td>
<td>7.59</td>
<td>0.50</td>
</tr>
<tr>
<td>Shoe</td>
<td>93.10</td>
<td>9.10</td>
<td>2.18</td>
</tr>
<tr>
<td>Brand</td>
<td>59.27</td>
<td>33.91</td>
<td>25.48</td>
</tr>
</tbody>
</table>

Time horizon fixed at the monthly level and geography aggregated to the CSA-State level. Brand and Brand-Category information only includes data for which brand and shoe category information are available.
Table 5: Demand Estimates with Adjusted Shares - Men’s

<table>
<thead>
<tr>
<th></th>
<th>Local FE Unadjusted (1)</th>
<th>Local FE Adjusted (2)</th>
<th>National FE Adjusted (3)</th>
<th>Local RE Adjusted (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>−0.004***</td>
<td>−0.003***</td>
<td>−0.012***</td>
<td>−0.016***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Comfort</td>
<td>0.025***</td>
<td>0.010***</td>
<td>0.012*</td>
<td>0.027***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Look</td>
<td>0.001</td>
<td>0.002</td>
<td>0.032***</td>
<td>0.080***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.009)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.062***</td>
<td>0.055***</td>
<td>0.185***</td>
<td>0.239***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.014)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>No Review</td>
<td>0.356***</td>
<td>0.281***</td>
<td>0.982***</td>
<td>1.497***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.077)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Intercept</td>
<td>−10.643***</td>
<td>−13.876***</td>
<td>−11.575***</td>
<td>−12.391***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.254)</td>
<td>(0.272)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>λ</td>
<td>0.562***</td>
<td>0.523***</td>
<td>0.633***</td>
<td>0.496***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.015)</td>
<td>(0.025)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>σ</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>*</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Category</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Color</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Month</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

N: 38.2mil 38.2mil 179,416 179,416

Zeroes: 95.03% → 0% 95.03% → 0% 8.23% → 0% 8.23% → 0%

Price Elast.

| Mean     | −0.936 | −0.784 | −3.599 | −3.002 |
| Std. Dev. | (0.528) | (0.582) | (2.673) | (2.526) |

Notes: Estimated at the monthly level. “Local FE” (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 FE” (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. 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

<table>
<thead>
<tr>
<th></th>
<th>Local FE Unadjusted (1)</th>
<th>Local FE Adjusted (2)</th>
<th>National FE Adjusted (3)</th>
<th>Local RE Adjusted (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>−0.009***</td>
<td>−0.002***</td>
<td>−0.006***</td>
<td>−0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Comfort</td>
<td>0.053***</td>
<td>0.016***</td>
<td>0.007</td>
<td>0.020***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Look</td>
<td>−0.020***</td>
<td>−0.007***</td>
<td>−0.008</td>
<td>−0.003</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.104***</td>
<td>0.047***</td>
<td>0.103***</td>
<td>0.158***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>No Review</td>
<td>0.532***</td>
<td>0.195***</td>
<td>0.500***</td>
<td>0.772***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.010)</td>
<td>(0.029)</td>
<td>(0.120)</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.165)</td>
<td>(0.112)</td>
<td>(0.159)</td>
</tr>
<tr>
<td>λ</td>
<td>0.279***</td>
<td>0.183***</td>
<td>0.654***</td>
<td>0.453***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>σ</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>*</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Brand</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Color</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Month</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N</td>
<td>82.6mil</td>
<td>82.6mil</td>
<td>387,745</td>
<td>382,747</td>
</tr>
<tr>
<td>Zeroes</td>
<td>95.17% → 0%</td>
<td>95.17% → 0%</td>
<td>9.41% → 0%</td>
<td>9.41% → 0%</td>
</tr>
<tr>
<td>Price Elast.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>−1.293</td>
<td>−0.250</td>
<td>−2.062</td>
<td>−1.547</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.848)</td>
<td>(0.220)</td>
<td>(1.814)</td>
<td>(1.360)</td>
</tr>
</tbody>
</table>

Notes: Estimated at the monthly level. “Local FE” (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 FE” (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. Standard errors in parentheses.

* estimates for across-market heterogeneity in specification (4) are in Table 7
Table 7: Parameter Estimates of Across-Market Heterogeneity: $\sigma_j = h(\cdot)$

<table>
<thead>
<tr>
<th></th>
<th>Men (1)</th>
<th>Women (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boat</td>
<td>0.758***</td>
<td>0.916***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Boots</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(6.638)</td>
<td>(1.163)</td>
</tr>
<tr>
<td>Clogs</td>
<td>1.419***</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.445)</td>
</tr>
<tr>
<td>Flats</td>
<td>–</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(4.008)</td>
<td></td>
</tr>
<tr>
<td>Heels</td>
<td>–</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(1.497)</td>
<td></td>
</tr>
<tr>
<td>Loafers</td>
<td>0.009</td>
<td>0.321***</td>
</tr>
<tr>
<td></td>
<td>(0.761)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Oxfords</td>
<td>0.017</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.383)</td>
<td>(11.425)</td>
</tr>
<tr>
<td>Sandals</td>
<td>0.462***</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(1.264)</td>
</tr>
<tr>
<td>Slippers</td>
<td>1.066***</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Sneakers</td>
<td>0.555***</td>
<td>0.904***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.011)</td>
</tr>
</tbody>
</table>

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: Demand Estimates with Empirical Shares

<table>
<thead>
<tr>
<th>Category</th>
<th>Men Local (1)</th>
<th>Men National (2)</th>
<th>Women Local (3)</th>
<th>Women National (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-0.004***</td>
<td>-0.010***</td>
<td>-0.009***</td>
<td>-0.016***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Comfort</td>
<td>0.025***</td>
<td>0.027***</td>
<td>0.053***</td>
<td>0.065***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.002)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Look</td>
<td>0.001</td>
<td>0.031***</td>
<td>-0.020***</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.062***</td>
<td>0.134***</td>
<td>0.104***</td>
<td>0.232***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.009)</td>
<td>(0.003)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>No Review</td>
<td>0.356***</td>
<td>0.789***</td>
<td>0.532***</td>
<td>1.311***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.051)</td>
<td>(0.013)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-10.643***</td>
<td>-11.182***</td>
<td>-12.490***</td>
<td>-15.008***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.199)</td>
<td>(0.039)</td>
<td>(0.224)</td>
</tr>
<tr>
<td>λ</td>
<td>0.562***</td>
<td>0.658***</td>
<td>0.279***</td>
<td>0.287***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.019)</td>
<td>(0.006)</td>
<td>(0.022)</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Category</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

N: 1,899,674 164,647 3,989,579 351,259
Zeros: 95.03% 8.23% 95.17% 9.41%
Price Elast. Mean: -0.936 -3.139 -1.293 -2.741
Price Elast. Std. Dev.: (0.528) (2.193) (0.848) (2.213)

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 9: Demand Estimates Market-by-Market

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical Shares</td>
<td>Adjusted Shares</td>
<td></td>
<td>Empirical Shares</td>
<td>Adjusted Shares</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>−0.004</td>
<td>−0.002</td>
<td></td>
<td>−0.003</td>
<td>−0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−0.150, 0.039]</td>
<td>[−0.010, 0.000]</td>
<td></td>
<td>[−0.033, 0.017]</td>
<td>[−0.004, 0.000]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.0%**</td>
<td>82.6%**</td>
<td></td>
<td>36.2%**</td>
<td>93.9%**</td>
<td></td>
</tr>
<tr>
<td>Comfort</td>
<td>0.022</td>
<td>0.010</td>
<td>0.012</td>
<td>0.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−0.588, 2.852]</td>
<td>[−0.009, 0.034]</td>
<td></td>
<td>[−0.436, 0.424]</td>
<td>[−0.025, 0.045]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.1%**</td>
<td>61.0%**</td>
<td></td>
<td>17.4%**</td>
<td>70.9%**</td>
<td></td>
</tr>
<tr>
<td>Look</td>
<td>0.066</td>
<td>−0.003</td>
<td>0.014</td>
<td>−0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−1.153, 5.155]</td>
<td>[−0.030, 0.034]</td>
<td></td>
<td>[−0.450, 1.914]</td>
<td>[−0.047, 0.021]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.0%**</td>
<td>33.8%**</td>
<td></td>
<td>10.8%**</td>
<td>46.0%**</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>−0.007</td>
<td>0.040</td>
<td>0.033</td>
<td>0.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−5.667, 1.623]</td>
<td>[−0.002, 0.175]</td>
<td></td>
<td>[−0.484, 0.469]</td>
<td>[−0.000, 0.086]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.2%**</td>
<td>90.6%**</td>
<td></td>
<td>28.2%**</td>
<td>94.8%**</td>
<td></td>
</tr>
<tr>
<td>No Review</td>
<td>0.402</td>
<td>0.185</td>
<td>0.239</td>
<td>0.135</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−2.534, 16.907]</td>
<td>[−0.022, 0.826]</td>
<td></td>
<td>[−1.262, 8.215]</td>
<td>[−0.036, 0.364]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.9%**</td>
<td>79.8%**</td>
<td></td>
<td>31.5%**</td>
<td>92.5%**</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−10.345</td>
<td>−15.602</td>
<td>−12.742</td>
<td>−20.479</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−70.830, 159.316]</td>
<td>[−51.493, −3.872]</td>
<td></td>
<td>[−19.807, −7.695]</td>
<td>[−60.767, −8.640]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>74.2%**</td>
<td>100.0%**</td>
<td></td>
<td>99.5%**</td>
<td>100.0%**</td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td>0.557</td>
<td>0.444</td>
<td>0.069</td>
<td>0.160</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[−3.848, 15.910]</td>
<td>[−0.170, 1.032]</td>
<td></td>
<td>[−1.440, 1.161]</td>
<td>[−0.102, 0.878]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.8%**</td>
<td>87.8%**</td>
<td></td>
<td>18.8%**</td>
<td>63.8%**</td>
<td></td>
</tr>
</tbody>
</table>

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.
### Table 10: Gains From Increasing Variety

<table>
<thead>
<tr>
<th>Assortment</th>
<th>Surplus Increase</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Local FE</td>
<td>Local RE</td>
</tr>
<tr>
<td>Localized</td>
<td>$mil</td>
<td>0.2</td>
<td>16.8</td>
</tr>
<tr>
<td>CS</td>
<td>%</td>
<td>0.0%</td>
<td>9.9%</td>
</tr>
<tr>
<td>National</td>
<td>$mil</td>
<td>64.4</td>
<td>23.3</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>9.5%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Localized</td>
<td>$mil</td>
<td>0.1</td>
<td>29.1</td>
</tr>
<tr>
<td>Rev.</td>
<td>%</td>
<td>0.0%</td>
<td>14.2%</td>
</tr>
<tr>
<td>National</td>
<td>$mil</td>
<td>33.1</td>
<td>38.1</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>12.8%</td>
<td>19.4%</td>
</tr>
</tbody>
</table>

Results based on parameter estimates in Table 5 and Table 6. Local 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.

### Table 11: Average Revenue Share of Products Outside of the Top 3,000

<table>
<thead>
<tr>
<th>Market</th>
<th>Small Sample Data</th>
<th>Model</th>
<th>Large Sample Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National</td>
<td>27.7 %</td>
<td>26.5 %</td>
<td>27.6 %</td>
</tr>
<tr>
<td>Local</td>
<td>0.4 %</td>
<td>0.1 %</td>
<td>19.9 %</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National</td>
<td>44.8 %</td>
<td>44.5 %</td>
<td>45.2 %</td>
</tr>
<tr>
<td>Local</td>
<td>1.9 %</td>
<td>1.6 %</td>
<td>40.4 %</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National</td>
<td>53.7 %</td>
<td>53.4 %</td>
<td>53.8 %</td>
</tr>
<tr>
<td>Local</td>
<td>3.6 %</td>
<td>3.2 %</td>
<td>47.0 %</td>
</tr>
</tbody>
</table>

Results based on the Local RE parameter estimates in Table 5 and Table 6. The small sample simulation of the model holds fixed the number of purchases at each location to that observed in the data. The large sample simulation of model is calculated using the simulated local shares, effectively assuming a continuum of consumers at each location.
Table 12: Robustness: Overstatement of Consumer Welfare Increase

<table>
<thead>
<tr>
<th>Assortment Size</th>
<th>Percent Increase</th>
<th>Absolute Increase ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loc.</td>
<td>Nat.</td>
</tr>
<tr>
<td>Baseline</td>
<td>14.3</td>
<td>17.3</td>
</tr>
<tr>
<td>Threshold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Baseline</td>
<td>24.5</td>
<td>30.2</td>
</tr>
<tr>
<td>3,000</td>
<td>29.0</td>
<td>35.2</td>
</tr>
<tr>
<td>6,000</td>
<td>14.3</td>
<td>17.2</td>
</tr>
<tr>
<td>12,000</td>
<td>4.6</td>
<td>5.6</td>
</tr>
<tr>
<td>24,000</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

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 $\log(a_t) = \beta_0 + \beta_1 \log(p_t) + \epsilon_t$, 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 13: Robustness: Retail Revenue

<table>
<thead>
<tr>
<th>Assortment Size</th>
<th>Percent Increase</th>
<th>Absolute Increase ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loc.</td>
<td>Nat.</td>
</tr>
<tr>
<td>Baseline</td>
<td>14.5</td>
<td>17.2</td>
</tr>
<tr>
<td>Threshold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Baseline</td>
<td>25.4</td>
<td>30.5</td>
</tr>
<tr>
<td>3,000</td>
<td>27.5</td>
<td>32.6</td>
</tr>
<tr>
<td>6,000</td>
<td>14.6</td>
<td>17.3</td>
</tr>
<tr>
<td>12,000</td>
<td>5.2</td>
<td>6.3</td>
</tr>
<tr>
<td>24,000</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

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 $\log(a_t) = \beta_0 + \beta_1 \log(p_t) + \epsilon_t$, 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

\[ s_{jp}^\prime = \frac{M \cdot s_j + 1}{M + J + 1}. \]

Note using that \( s_{jp}^\prime \) results in a consistent estimator of \( \delta \) as the market size \( M \to \infty \) as long as \( s_j \to \pi_j \). 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 \( \beta \) estimator.”

The key is to back out the conditional distribution of choice probabilities, \( \pi_t \), 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 \((\pi, M)\);
2. \( F_{\pi|M,J} \): the conditional distribution of \( \pi \) given \((M, J)\).

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

\[
M \cdot s \sim MN(\pi,M).
\]  \( \text{(C.1)} \)

\( F_{\pi|M,J} \) 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

\[
\frac{\pi}{(1 - \pi_0)} \sim Dir(\varnothing 1),
\]  \( \text{(C.2)} \)

for an unknown parameter \( \varnothing \).

Equations C.1 and C.2 then imply

\[
\frac{s}{(1 - s_0)} \sim DCM(\varnothing 1, M(1 - s_0)),
\]  \( \text{60} \)
where $DCM(\cdot)$ denotes a Dirichlet compound multinomial distribution. $\theta$ can 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 $\pi$, with a Dirichlet prior and multinomial likelihood,

$$F \sim Dir(\theta + M \cdot s).$$

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

$$E[\log(x_j)] = \psi(\theta) - \psi(\theta' 1_d),$$

Thus,

$$E\left[\log\left(\frac{\pi_j}{1 - s_0}\right)\right] = E[\log(\pi_j)] - E[\log(1 - s_0)]$$

$$= \psi(\theta + M \cdot s_j) - \psi((\theta + M \cdot s)' 1_d),$$

which implies

$$\hat{\delta} = \log(\hat{\pi}_j) - \log(\hat{\pi}_0) = E[\log(\pi_j)] - E[\log(\pi_0)]$$

$$= \psi(\theta + M \cdot s_j) - \psi(M \cdot s_0).$$

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

$$\log(\hat{\pi}_j) - \log(\hat{\pi}_c) = E[\log(\pi_j)] - E[\log(\pi_c)]$$

$$= \psi(\theta + M \cdot s_j) - \psi\left(\sum_{j \in c} \theta + M \cdot s_j\right).$$
D 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

\[ u_{i\ell j} = \delta_j + \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \xi_j + \eta_{\ell j} + \zeta_{ic} + (1 - \lambda)\epsilon_{i\ell j} \]

\[ = -4 + -.75 x_{1j} + .75 x_{2j} + \xi_j + \eta_{\ell j} + \zeta_{ic} + (1 - .5)\epsilon_{i\ell j} \]

The normalized outside good gives utility \( u_{i\ell 0} = \zeta_{i0} + (1 - \lambda)\epsilon_{i\ell 0} \). Here we assume both characteristics are exogenous from the unobservable \( \xi_j \); however, given real data, instrumental variables can be used on these characteristics. We assign distributions on the data generating process according to Table 14 below.

Table 14: Data generating process for Monte Carlo study

<table>
<thead>
<tr>
<th>Definition</th>
<th>Variable</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic 1</td>
<td>( x_1 )</td>
<td>( N(0, 1) )</td>
</tr>
<tr>
<td>Characteristic 2</td>
<td>( x_2 )</td>
<td>( N(0, 1.5^2) )</td>
</tr>
<tr>
<td>National Unobservable</td>
<td>( \xi )</td>
<td>( N(0, 1) )</td>
</tr>
<tr>
<td>Local Unobservable</td>
<td>( \eta )</td>
<td>( N(0, \sigma_c = 1) )</td>
</tr>
<tr>
<td>Individual Unobservable</td>
<td>( \epsilon )</td>
<td>T1EV</td>
</tr>
<tr>
<td>Local-Category Product Size</td>
<td>( J_c )</td>
<td>175</td>
</tr>
<tr>
<td>Num. of Categories</td>
<td>( C )</td>
<td>3</td>
</tr>
<tr>
<td>Num. of Periods</td>
<td>( T )</td>
<td>10</td>
</tr>
<tr>
<td>Market Population</td>
<td>( M )</td>
<td>2000000</td>
</tr>
<tr>
<td>Num. of Local Markets</td>
<td>( L )</td>
<td>200</td>
</tr>
<tr>
<td>Population Distribution</td>
<td>( \omega_\ell )</td>
<td>1/L</td>
</tr>
</tbody>
</table>

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 \( \sigma, \lambda \),

1. Recover \( \delta_j^{(k)} \) using the inversion (Equation 2.4),
2. Given, $\delta^{(k)}_j$, calculate GMM objective using micro moments and orthogonality conditions on $\xi_j^{(k)}, G()$.

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

4. Given parameter estimates $\hat{\theta}_1$, calculate the weighting matrix

$$\hat{W} = \left(G(\hat{\theta}_1; Z) G(\hat{\theta}_1; Z)^T \right)^{-1},$$

5. With $\hat{W}$ repeat steps 0-3 to obtain $\hat{\theta}_2$, 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 $k$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Bias</th>
<th>MSE</th>
<th>Reject. Rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-4</td>
<td>0.065</td>
<td>0.998</td>
<td>6.122</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.75</td>
<td>-0.016</td>
<td>0.022</td>
<td>6.122</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.75</td>
<td>0.015</td>
<td>0.022</td>
<td>6.122</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>0.053</td>
<td>0.044</td>
<td>5.102</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>1</td>
<td>0.052</td>
<td>0.043</td>
<td>5.102</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>1</td>
<td>0.054</td>
<td>0.043</td>
<td>5.102</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.5</td>
<td>0.074</td>
<td>0.018</td>
<td>6.122</td>
</tr>
</tbody>
</table>

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

Table 15 presents the results for our Monte Carlo exercises, using 100 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.
Figure 10: Histogram of Monte Carlo parameter estimates