Demand Heterogeneity: Implications for Welfare Estimates and Policy

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Thomas W. Quan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Thomas J. Holmes and Amil Petrin, Advisors

July, 2015

Acknowledgements

I am deeply indebted to my advisors, Thomas J. Holmes and Amil Petrin, and my committee members, Joel Waldfogel and Kyoo il Kim, for their continued guidance and encouragement. Their constructive evaluation and insights have been invaluable.

Additionally, this research has benefited from conversations with Brian Adams, Kailin Clarke, Amit Gandhi, Ellen McGrattan, Keaton Miller, Imke Reimers, Boyoung Seo, Ben Shiller, Ethan Singer, Matt Shapiro, Catherine Tucker, Chunying Xie, members of the University of Minnesota Applied Micro Workshop, and participants of conferences and seminars I have presented at while attending the University of Minnesota. Special thanks to Kevin R. Williams who co-authored the first two chapters of this dissertation.

I acknowledge the financial support of Richard and Ellen Sandor and the Graduate School of the University of Minnesota. I also acknowledge support from the Minnesota Supercomputing Institute for providing computational resources. Finally, I would like to thank the Department of Economics staff, especially Caty Bach, Kara Kersteter, Sara Shuford, and Wendy Williamson.

Dedication

To my family and friends. In memory of my grandmother, Sau K. Goon, and aunt, Susie Goon.

Abstract

This dissertation is comprised of three essays, the first two of which are coauthored with Kevin R. Williams. In the first chapter, we develop new empirical methodology to estimate consumer demand and across-market consumer demand heterogeneity when faced with products that have zero sales at the local market level, but have positive sales at the aggregate (national) market level. The absence of sales is a common problem among data sets with a large number of products and this problem is exacerbated when markets are narrowly defined. Observing zero sales creates two major empirical issues. First, standard estimation techniques rely on the Law of Large Numbers in sales. When the Law of Large Numbers does not apply, mechanically forcing the estimation will result in biased demand estimates. Second, failing to account for small sample sizes may create spurious heterogeneity. That is, the randomness generated by small sample sizes may overstate the degree to which demand differs across locations. Alternatively, aggregating over markets to obtain a large sample size will obscure the heterogeneity across local markets. We propose a modification to Berry (1994) and Berry, Levinsohn, Pakes (1995), where both local and aggregate level sales information are used to recover geographically varying demand. This new estimation approach is easy to implement and we will show, using Monte Carlos exercises, that it fits the data well.

In the second chapter, we apply this new methodology to examine the welfare benefit of the increased access to variety from online retail. The proliferation of online retail has greatly increased consumers' access to variety. The value of this additional variety depends crucially on the extent to which local demand can be captured by local retailers. The existing literature has found huge welfare benefits from online variety, but these studies have been limited by national level data. As a result, they are unable to speak to the differences in demand across locations. Using an original data set of online shoe sales, we show that failing to account for across-market heterogeneity can greatly overstate the consumer welfare gain from increases in product variety.

In the third chapter, I develop a theory of non-collusive basing-point pricing under Bertrand competition. Basing-point pricing occurs when the delivered prices faced by a consumer is determined by the consumer's distance from a common location, known as the basing point. That is, the price faced by the consumer is equal to the price charged by firms at the basing point plus the cost of transportation from the basing point to the consumer's location regardless of the actual location of the selling firm. The existing non-collusive theories of basing-point pricing rely on the exploitation of market power by non-basing point firms to obtain the basing-point result. I show that basing-point pricing is consistent with Bertrand competition by non-basing point firms when these firms face increasing marginal costs of production or capacity constraints.

Contents

List of Tables

List of Figures

Chapter 1

Estimating Across-Market Demand Heterogeneity in Discrete Choice Models with a Sparsity of Local Sales

1.1 Introduction

Modern techniques for estimating differentiated product demand systems (discrete choice), such as Berry (1994) and Berry, Levinsohn, and Pakes (1995), have been developed to back out demand parameters from aggregate level data. While extensions to this literature have allowed for the incorporation of micro-level data, the underlying theory behind these models assumes the existence of a continuum of consumers, resulting in every product having a strictly positive probability of purchase. Empirical applications must then rely on data in which a large number of purchases are observed and aggregated to compute a "market" share for each product. These market shares are the empirical analogues of the theoretical purchase probabilities and, in the estimation, market shares are assumed to be equal to the true underlying choice probabilities.

Unfortunately, real-world data are often not as pristine as the theory demands. When the number of purchases are not sufficiently large, observed market shares will be poor estimates of the true underlying choice probabilities. This small samples problem is most obvious when some products are observed with zero sales. These zeros are problematic for standard demand estimation strategies because they create selection bias in the estimates, resulting in demand elasticities that are biased toward zero (Berry, Linton, and Pakes 2004, Gandhi, Lu, and Shi 2013, Gandhi, Lu, and Shi 2014), i.e. consumers are estimated to be too inelastic. Gandhi, Lu, and Shi (2014) shows that the biases can be quite severe even with a relatively small fraction of zeros in the data.

While zeros may have been a minor nuisance in the past, researchers are now frequently encountering a large number of products with zero market shares in their data due to technical advances in data collection and the increasing availability of extremely detailed, disaggregated sales data. This places critical emphasis on the definition of a market. Since increasing levels of aggregation will be less likely to exhibit a small samples problem, researchers using this type of data are faced with a trade-off between obtaining consistent demand estimates and performing analysis at narrow levels of disaggregation.

This chapter focuses on situations in which researchers have geographically disaggregated sales data that exhibits a small sample problem at the "local" market (i.e. city) level, but has a sufficiently large sample at the "aggregate" market (i.e. national) level. Simply aggregating over local markets may not be satisfactory, however, because, as highlighted in recent literature, many products face a high degree of heterogeneity in demand across markets[1] and aggregation would obscure these differences. Not only is understanding this heterogeneity interesting in its own right, accounting for these differences in demand may be important to correctly modeling substitution and welfare effects at the aggregate level.

We propose a technique to obtain consistent estimates of the demand parameters, while retaining information on the across-market demand heterogeneity. Unlike the standard approaches, which would use the disaggregated local market shares directly to identify a product-market level fixed effect, we instead use the local market share

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

information to form a set of micro-moments that augment the aggregated sales data (Petrin 2002). The differences in a product's local market shares allow us to identify the variance of product-market level random effects. In this way, we can allow for estimated substitution patterns and welfare to reflect differences in the demand for products across locations. The trade-off we make to obtain consistent demand estimates is that our approach estimates the distribution of heterogeneity, rather than the exact realization.

Our estimation strategy exploits the structure of the model to, in a sense, separate the problem into two parts. At the aggregate level, our approach effectively mimics the standard approach and we are able to pin down the mean parameters that are common across all local markets. Separately, our micro-moments are used to estimate the distribution of consumer heterogeneity across markets, while explicitly accounting for small samples.[2] If we failed to address the small samples, we would overstate the degree of heterogeneity across markets. In an influential paper, Ellison and Glaeser (1997) argue that with only a small number of establishments in an industry, naive calculations will overstate the differences in suitability across locations for the industry. The same point applies when evaluating differences in demand across locations, small samples may lead to inferring a level of across-market demand heterogeneity that is spurious. In the context of discrete choice demand models, the result will be to overstate demand for products with observed sales. This, in turn, overstates the consumer welfare generated by these products.[3]

After presenting the models and our estimation strategy, we perform a series of Monte Carlo exercises to examine the performance of our estimation technique relative to existing methodologies. We find that our estimation strategy fits the data well, addressing biases in the estimated demand parameters created by local level small sample sizes, while retaining the across-market demand heterogeneity of interest. We also show that retaining this heterogeneity across-markets is critical to computing estimates of

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

[3] Note that taking seriously zero observed sales when the true choice probability is greater than zero will artificially increase the choice probability of products with observed purchases. Thus, products with observed purchases will have inflated mean utilities, and hence, consumer welfare will be overstated.

consumer welfare. As expected, ignoring the small samples problem in the local sales data will overstate consumer welfare by overstating the differences across markets. On the other hand, aggregation will understate consumer welfare by obscuring the differences across markets.[4]

The rest of the chapter will be organized as follows. Section 2 introduces our simple logit and nested logit models incorporating across-market demand heterogeneity and Section 3 discusses the estimation of these models. In Section 4, we conduct a series of Monte Carlo exercises to show that our estimation performs well. Finally, Section 5 concludes the chapter.

1.2 Model

Each consumer solves a discrete choice utility maximization problem: Consumer $i \in I_\ell$ in location $\ell \in L$ will purchase a product j if and only if the utility derived from product j is greater than the utility derived from any other product, $u_{i\ell j} \geq u_{i\ell j'}, \forall j' \in J \cup \{0\}$. For a product $j \in J \cup \{0\}$, the utility of a consumer $i \in I_\ell$ in location $\ell \in L$ is given by

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

where δ_j is the mean utility of product j for the aggregate (national) population of consumers and $\nu_{i\ell j}$ is a random utility component that is heterogeneous across consumers and locations. This chapter considers two forms of the random utility component corresponding to the simple logit and nested logit discrete choice demand models.

1.2.1 Simple Logit

In our simple logit model, the random utility component can be decomposed into

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

where $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution. These terms decompose the heterogeneity in the random utility among consumers into an "across-market" effect, $\eta_{\ell j}$, and a "within-market" effect, $\varepsilon_{i\ell j}$. When $\eta_{\ell j} = 0$ for all $\ell \in L, j \in J$,

[4] Conversely, ignoring small samples will understate gains to variety, while obscuring across market heterogeneity will overstate gains to variety.

then the model reduces to a standard "love of variety" logit model, where there is no distinction between local and aggregate level preferences. That is, all heterogeneity is within-market heterogeneity, which is identical across locations.

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

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

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

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

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

1.2.2 Nested Logit

In the nested logit model, products are categorized (or grouped) into mutually exclusive and exhaustive sets $c \in C \cup \{0\}$.[5] The random utility component can be decomposed into

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

where $\eta_{\ell c}$ is a location-category fixed effect, $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution and ζ_{ic} has the unique distribution such that $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$ has a Type-1 extreme value distribution. For a consumer i, ζ_{ic} is common to all products in the same category and has a distribution that depends on the nesting parameter λ, $0 \leq \lambda < 1$. λ determines the within-category correlation of utilities. When $\lambda \to 1$ consumers will only substitute to products within the same group and when $\lambda = 0$ the model simplifies to the simple logit case. As in the simple logit model, these terms decompose the heterogeneity in the random utility among consumers into an "across-market" effect, $\eta_{\ell c} + \eta_{\ell j}$, and a "within-market" effect, $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$. When $\eta_{\ell c} =$

[5] Our methodology can be extended to an arbitrary number of subnests. For details see Appendix A.1.

$\eta_{\ell j} = 0$ for all $\ell \in L, c \in C, j \in J$, then the model reduces to a standard nested logit model, where there is no distinction between local and aggregate level preferences.

For any fixed location $\ell \in L$, characterized by $\eta_\ell = \left\{\{\eta_{\ell c}\}_{c=1}^{C}, \{\eta_{\ell j}\}_{j=1}^{J}\right\}$, we can integrate out over the within-market heterogeneity, $\varepsilon_{i\ell j}$. Since $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$ is distributed Type-1 extreme value, integrating over them forms location-specific consumer choice probabilities,

$$\pi_{\ell j} = \pi_j(\eta_\ell; \delta, \lambda) = \frac{\left(\sum_{j \in c} \exp\{\delta_{\ell j}/(1-\lambda)\}\right)^{1-\lambda}}{1 + \sum_{c' \in C}\left(\sum_{j' \in c'} \exp\{\delta_{\ell j'}/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{\exp\{\delta_{\ell j}/(1-\lambda)\}}{\sum_{j' \in c} \exp\{\delta_{\ell j'}/(1-\lambda)\}},$$

where

$$\delta_{\ell j} = \delta_j + \eta_{\ell c} + \eta_{\ell j}.$$

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

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

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

The key difficulty in both setups is that the exact location-specific effects $\eta_{\ell j}$ cannot be recovered from the sales data because of the sparsity of sales within disaggregated locations. To circumvent this problem, we use a random-effects specification, where $\eta_{\ell j}$ is drawn independently from a normal distribution, $N(0, \sigma_j^2)$. The relative importance of the across-market component is then determined by σ_j^2. In the next section, we outline a procedure that incorporates micro-moments – moments generated from the disaggregated local shares – to estimate the distribution of $\eta_{\ell j}$. This allows us to integrate out (or aggregate) over $\eta_{\ell j}$ and obtain δ_j. We can then use traditional estimation techniques to back out the parameters in δ_j. Crucially, our procedure accounts for the fact that local market share observations suffer from small samples.

1.3 Estimation

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

$$\pi_j \approx \begin{cases} \sum_{\ell=1}^{L} \omega_\ell \pi_j(\tilde{\eta}_\ell; \delta) & \text{Simple Logit} \\ \sum_{\ell=1}^{L} \omega_\ell \pi_j(\tilde{\eta}_\ell; \delta, \lambda) & \text{Nested Logit} \end{cases},$$

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

$$\pi_j = \begin{cases} \pi_j(\delta; \sigma), \ j = 1, ..., J, & \text{Simple Logit} \\ \pi_j(\delta, \lambda; \sigma), \ j = 1, ..., J, & \text{Nested Logit} \end{cases},$$

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

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

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

Following Berry, Levinsohn, and Pakes (1995), for a fixed σ, we can use linear instrumental variables z_j, such that $E[z_j \xi_j] = 0$ and $E[z_j'(p_j, x_j)]$ has full rank, to identify (α, β) as a function of σ. However, the existing instruments used in the literature[6] typically provide little to no identifying power for the non-linear parameter σ (Gandhi and Houde 2014). Instead we use the disaggregated information in our data to augment the instrumental variable conditions with an additional set of micro-moments that provide direct information on σ (Petrin 2002).

[6] For example, BLP instruments

1.3.1 Micro-Moments

Let $P0_{\ell j}(\sigma)$ be the probability that a product j has zero sales given N_ℓ purchases are observed in location ℓ. We then define,

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

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

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

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

$$m(\sigma) = \frac{1}{J}\sum_{j=1}^{J}\left(P0_j(\sigma) - \hat{P0}_j\right). \tag{1.1}$$

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

1.3.2 Inverting the Market Share

In this subsection, we show that the inverse of our market shares takes a convenient analytical form, which will simplify the simulation of our local choice probabilities. While small sample sizes make the observed local market shares unreliable estimates of the underlying choice probabilities for individual products, we assume the local choice probability of the outside good, $\pi_{\ell 0}$, and the local choice probabilities of each category, in the nested logit case, $\pi_{\ell c}$, are well estimated by the data. We present our market share inversions in the following two propositions:

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

$$\delta_j = \log \pi_j - \log \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell 0} \exp\{\eta_{\ell j}\}. \tag{1.2}$$

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

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

Aggregated choice probabilities are then

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

Next, define

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

so that $\pi_j = \sum_{\ell=1}^{L} \omega_\ell (1-\pi_{\ell 0}) \frac{\exp\{\delta_j + \eta_{\ell j}\}}{D_\ell}$. We normalize the utility of the outside good – both in terms of product characteristics, as well as, the unobserved taste preference across locations. This means the probability of choosing the outside good at location ℓ is equal to

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

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

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

Finally, taking logs, we then have

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

or

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

■

Proposition 2 (Nested Logit)**.** *For any set of* $\{\eta_\ell\}_{\ell=1}^{L}$*, the market share inversion takes the following analytic form,* $\forall j \in J$*,*

$$\delta_j = (1-\lambda)\left(\log(\pi_j) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c} \left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda}\right\}\right)\right). \tag{1.3}$$

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

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

Aggregated choice probabilities are then

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

Next, define

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

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

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

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

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

thus

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

Plugging into the aggregate choice probabilities, gives

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

Finally, taking logs we have our inversion:

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

or

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

■

Since the population shares, ω_ℓ, the outside good shares, $\pi_{\ell 0}$, and, in the case of nested logit, the category shares, $\pi_{\ell c}$ are assumed to be known, the inversions relate δ_j to the aggregated share data, π_j. Additionally, notice that this reduces to the standard Berry (1994) simple and nested logit inversions when $\eta_{\ell j} = 0$ for all $\ell \in L, j \in J$. In the next subsection, we describe how to estimate the distribution of heterogeneity using our micro-moments. We can then integrate out this distribution to obtain the mean utilities, δ_j, from the data, π_j, and proceed with traditional estimation techniques at the aggregate level.

1.3.3 Estimation Procedure

Local level mean utilities can be written as

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

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

$$\hat{\pi}_{\ell j} = \begin{cases} (1-\pi_{\ell 0})\frac{\delta_j+\sigma_j\bar{\eta}_{\ell j}}{\sum_{j'=1}^{J}\delta_{j'}+\sigma_{j'}\bar{\eta}_{\ell j'}} & \text{Simple Logit} \\ \pi_{\ell c}\frac{\exp\{(\delta_j+\sigma_j\bar{\eta}_{\ell j})/(1-\lambda)\}}{\sum_{j'\in c}\exp\{(\delta_{j'}+\sigma_{j'}\bar{\eta}_{\ell j'})/(1-\lambda)\}} & \text{Nested Logit} \end{cases}.$$

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

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

i.e. the probability we observe N_ℓ sales at location ℓ, none of which are of good j. This explicitly accounts for the small sample sizes at the location level. We then estimate $\hat{\sigma}$ such that it minimizes $m(\sigma)$ (Equation 1.1).

After obtaining $\hat{\sigma}$, the structure we have placed on the $\eta_{\ell j}$'s allows us to integrate them out by subtracting the sum of local level random effects according to Equations 1.2 and 1.3, for simple and nested logit, respectively.[7] Also note that, in the nested logit case, fixing $\pi_{\ell c}$ allows us to difference out the effect of the location level category fixed effect, $\eta_{\ell c}$.[8] We then estimate

$$\delta_j(\pi,\hat{\sigma}) = x_j\beta - \alpha p_j + \xi_j,$$

using standard instrumental variables methods to control for price endogeneity.

Identifying the Nesting Parameter

Nested logit provides the additional complication of identifying the nesting parameter, λ. In the Berry (1994) nested logit inversion, within category shares are correlated with the unobserved product quality creating an endogeneity problem. A similar issue arises in our inversion. Note that, with δ as defined in Equation 1.3,

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

[7] Since we take many draws over the distribution of $\eta_{\ell j}$, we can estimate (simulate to integrate) the summations in Equations 1.2 and 1.3 without explicitly knowing each individual $\eta_{\ell j}$

[8] This is akin to the demeaning (or within transformation) approach to fixed effects.

because ξ_j enters the aggregate product share, π_j, and the local level category shares, $\pi_{\ell c}$. Berry (1994) solves this problem by employing an instrument, $z_{j|c}$, that is correlated with the within category share, but uncorrelated with the unobserved product quality.[9] The same instrument can be employed here, since $z_{j|c}$ is correlated $\frac{\partial \delta_j(\pi,\sigma,\lambda)}{\partial \lambda}$ 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 estimation.

1.3.4 Identification

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

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

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

Parameters within δ_j are identified through the standard channels. In the cross-section through variation in aggregate sales given characteristics, (x_j, p_j), and across time periods through time varying characteristics and variation in the choice set J. For the nested logit model, identification of the nesting parameter, λ, is driven by changes to the category shares as the utilities of products within the category change or the number of products within the category changes.

1.4 Monte Carlo Exercises

In this section, we conduct a series of Monte Carlo exercises to examine the performance of our estimation procedure, where local market shares are used to estimate parameters governing across-market heterogeneity and aggregate shares are used to estimate parameters constant across markets. Parameters governing the across-market heterogeneity and mean aggregate demand are estimated jointly using Generalized Method of Moments (GMM). For comparison, we also present the results for the estimation using the Berry (1994) inversions on local and aggregate level data.

1.4.1 Simple Logit

We start by assigning parameters and drawing consumer purchases from disaggregated local shares. The true model has the form

$$u_{i\ell j} = \underbrace{\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \xi_j}_{\delta_j} + \eta_{\ell j} + \varepsilon_{i\ell j}.$$

The outside good gives utility $u_{i\ell 0} = \varepsilon_{i\ell 0}$. We assign distributions on the data generating process according to Table 1.1.

With the synthetic data, we have a a matrix of local shares across locations and products, $\mathbf{s}_{LJ}$. Demand at each locality is obtained from simulating $\lfloor \omega_\ell M \rfloor$ consumer purchases (of $J \cup \{0\}$) according to probabilities $(\mathbf{s}_{\ell J}, 1 - \sum_J s_{\ell j})$. Our estimation stacks two sets of moments:

1. To identify the across-market heterogeneity, σ, the micro-moments we use are, for each product, the percentage of local markets with zero sales implied by the model

Table 1.1: Data generating process distributions (Simple Logit)

Variable	Distribution
x_1	$\mathcal{N}(0,1)$
x_2	$\mathcal{N}(0,1)$
ξ	$\mathcal{N}(0,1)$
η	$\mathcal{N}(0,\sigma=1)$
ε	T1EV
J	500
T	1
M	500000
L	100
ω_ℓ	$1/L$

and match this to the observed percentage of local markets in which a product has zero sales:

$$E\left[P0_j(\sigma) - \hat{P0}_j\right] = 0.$$

2. To identify the mean utility parameters within δ, we employ the standard linear moments:

$$E[x_{jk}(\delta_j(\sigma) - X\beta)] = E[x_{jk} \cdot \xi_j] = 0, \text{ for } k = 0, 1, 2$$

where k indexes the independent variable.

Tables 1.2 and 1.3 present the simple logit estimation results for our Monte Carlo exercises. Table 1.2 compares the Berry inversion estimated using local market level data to our estimator. Biased parameters are highlighted in bold. The intercept term, β_0, is adjusted to vary the percentage of products with local market shares equaling zero. As we can see, the estimated parameters using local level data are all severely attenuated toward zero with this bias being greater the greater the percentage of local zeros. Meanwhile, our estimation technique fits quite well and is unaffected by the percentage of local zeros.

Table 1.3 compares the Berry inversion estimated using aggregate level data to our estimator. Here the across-market demand heterogeneity, σ, is varied and each of these average about 81% local zeros. With the aggregate level data, the percentage of local

Table 1.2: Monte Carlo - Varying Local Zeros (Simple Logit)

	True Value	Berry Loc.	Quan and Williams	Berry Loc.	Quan and Williams	Berry Loc.	Quan and Williams
		$\beta_0 = -5.5$		$\beta_0 = -8.25$		$\beta_0 = -11$	
β_0		**-4.787**	-5.506	**-7.103**	-8.25	**-8.224**	-11.078
		(0.066)	(0.050)	(0.038)	(0.043)	(0.024)	(0.063)
β_1	-1	**-0.595**	-0.994	**-0.479**	-1.005	**-0.236**	-1.064
		(0.029)	(0.042)	(0.033)	(0.043)	(0.037)	(0.055)
β_2	0.5	**0.297**	0.499	**0.247**	0.512	**0.120**	0.523
		(0.031)	(0.047)	(0.031)	(0.047)	(0.031)	(0.056)
σ	1		0.997		0.989		0.986
			(0.012)		(0.011)		(0.020)
Local	Zeros	20%		46%		81%	

For the Monte Carlo, we simulate 100 synthetic data sets and use the two sets of moments outlined above. The intercept is varied to vary the percentage of local zeros. Bold font highlights biased estimates. Standard errors in parentheses.

zeros has little effect on the demand estimates because, at the aggregate level, a large sample of consumers is observed. However, the aggregated estimation obscures the heterogeneity across markets, that is the across-market heterogeneity will be averaged across markets. Since the heterogeneity across markets enters the share equations non-linearly, averaging will result in a biased intercept term. The greater the across-market demand heterogeneity the greater the bias in the estimated intercept.

In the following chapter, we examine the consumer welfare effects of increasing product variety. Here, using the Monte Carlo exercises above, we examine the consumer welfare estimates made by the preceding estimation techniques. Consider a world where consumers are constrained to a subset of the most popular products in their local market and index total consumer welfare to 1 when consumers have access to the universe of products. Figure 1.1 compares the consumer welfare estimated using our technique (vertical-dash) to the Berry inversion using national (dash) and local (dash-dot) sales and to the true values (solid) implied by the model.

Ignoring the small samples sizes using local level data will overstate consumer welfare by making local markets appear more different from each other than they are in reality. In particular, products with zero observed sales in a local market are assumed to have

Table 1.3: Monte Carlo - Varying Across-Market Heterogeneity (Simple Logit)

	True Value	Berry Agg.	Quan and Williams	Berry Agg.	Quan and Williams	Berry Agg.	Quan and Williams
		$\sigma = .5$		$\sigma = 1$		$\sigma = 1.5$	
β_0	-11	**-11.009**	-11.125	**-10.598**	-11.078	**-9.966**	-11.054
		(0.067)	(0.068)	(0.056)	(0.063)	(0.049)	(0.058)
β_1	-1	-1.067	-1.067	-1.065	-1.064	-1.052	-1.054
		(0.068)	(0.068)	(0.055)	(0.055)	(0.063)	(0.062)
β_2	0.5	0.532	0.532	0.524	0.523	0.518	0.518
		(0.055)	(0.055)	(0.056)	(0.056)	(0.059)	(0.06)
σ			0.484		0.986		1.497
			(0.026)		(0.02)		(0.026)

For the Monte Carlo, we simulate 100 synthetic data sets and use the two sets of moments outlined above. Across-market demand heterogeneity is varied. The local level data averages about 81% zeros. Bold font highlights biased estimates. Standard errors in parentheses.

zero demand at that location. Thus, relatively few products in each location are then required to satisfy local demand. On the other hand, failing to account for the across-market demand heterogeneity will understate consumer welfare when demand is actually correlated by location because, if tastes were equally distributed across locations, it would require relatively more products to satisfy demand. Our technique, by properly accounting for both the local small sample sizes and across-market heterogeneity, is able to fit the true underlying consumer welfare.

1.4.2 Nested Logit

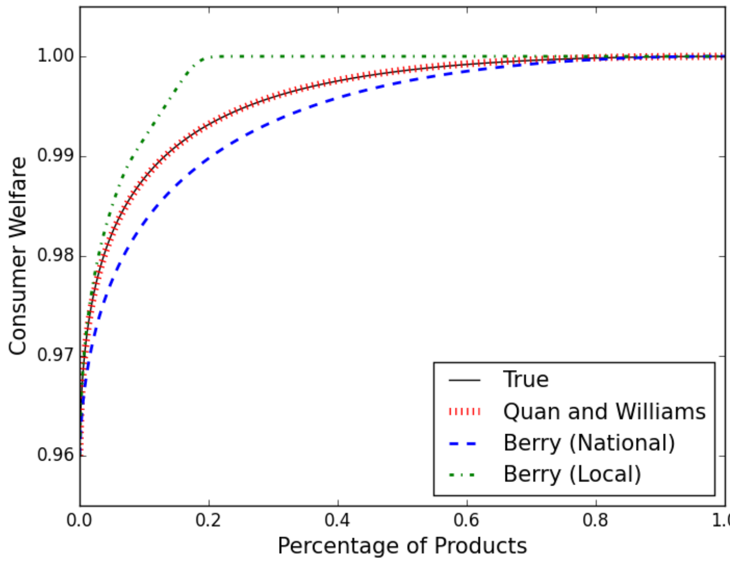
As in the simple logit case, we begin by assigning parameters and drawing consumer purchases from disaggregated local shares. The true model has the form

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

The outside good gives utility $u_{i\ell 0} = \zeta_{i0} + (1-\lambda)\varepsilon_{i\ell 0}$. We assign distributions on the data generating process according to Table 1.4.

Table 1.5 presents the nested logit estimation results for our Monte Carlo exercises. The results mirror the simple logit results, so we condense them into one table here. The

Figure 1.1: Consumer Welfare Estimates (Simple Logit)



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

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

Figure 1.2 compares the consumer welfare estimates under the nested logit model. As in the simple logit case, ignoring the small samples problem at the local level will overstate consumer welfare (dash-dot), while failing to account to the across-market demand heterogeneity will understate consumer welfare (dash). Again, our technique (vertical-dash), by properly accounting for both the local small sample sizes and across-market heterogeneity, is able to fit the true underlying consumer welfare (solid).

Table 1.4: Data generating process distributions (Nested Logit)

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

1.5 Conclusion

Researchers are increasingly faced with a trade-off between performing analysis at narrow levels of disaggregation and, due to small sample sizes at these disaggregated levels, obtaining consistent demand estimates. This chapter focused on situations in which researchers have geographically disaggregated sales data that exhibits a small sample problem at the local market (i.e. city) level, but has a large sample at the aggregate market (i.e. national) level.

We proposed an estimation technique to obtain consistent estimates, while retaining the across-market demand heterogeneity, for two commonly used discrete choice models, simple logit and nested logit.[10] Instead of using disaggregated local market shares directly, we use the local sales information to form a set of micro-moments to augment the aggregated sales data. Our estimation strategy exploits the structure of the model to separate the problem into two parts. At the aggregate level, our approach effectively mimics the standard approach and we are able to pin down the mean parameters that are common across all markets. Separately, our micro-moments are used to estimate the distribution of consumer heterogeneity across markets, while explicitly accounting for small samples.

[10] Future work will seek to extend this methodology further to the full random coefficients model.

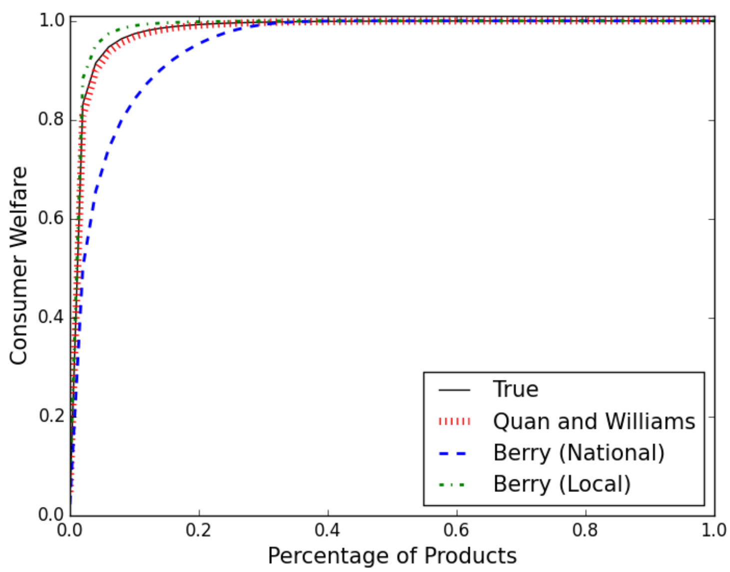
Table 1.5: Monte Carlo Demand Results (Nested Logit)

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

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

Finally, using a series of Monte Carlo exercises, we provided evidence that our estimation strategy fits the data well and highlight the importance of accounting for both local level small sample sizes and the across-market demand heterogeneity. Ignoring the small samples problem in the local sales data will overstate the heterogeneity in demand across markets leading to an overstatement of consumer welfare, while failing to account for differences across markets by aggregating over them will understate consumer welfare.

Figure 1.2: Consumer Welfare Estimates (Nested Logit)



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

Chapter 2

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

2.1 Introduction

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

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

We have collected an extremely detailed data set consisting of point-of-sale, product review, and inventory data from a large online retailer. One of the product categories the retailer sells is footwear, from which, we observe over 13.5 million shoe sales across more than 100,000 products. For each sale, we observe the date and time, shipping destination, price, and a wealth of information about the shoe. The richness of the data allows us define products and geographies at very narrow levels. For example, we are able to differentiate between the different colors of the same model of shoe and we can attribute sales to particular metro-areas. In addition, we collected data on shoe assortments across stores for a few large retail chains. This data provides us with direct evidence that firms are responding to across-market heterogeneity, as product assortments vary significantly across stores.[1]

Using our transactions data, we document large differences in demand for specific products across geographic markets. Since prices, product characteristics, and choice sets are the same for all geographic markets, these differences can only be rationalized by differences in local demand. To highlight the extent of these differences, consider the top 1,000 products at the metro level. On average, these products make up 87% of a metro area's total sales. Now consider the national sales of these same 1,000 products. These products only make up 12% of total national sales. These large differences suggest that even among top products, demand varies significantly across locations. To formally test for differences across markets, we use simple multinomial tests that compare local market shares to national market shares. These tests overwhelmingly reject the null hypothesis that consumers across markets have the same demand over shoes.

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

After showing that the data is inconsistent with a model devoid of across-market demand heterogeneity, we turn to estimating the gains from online variety. Our modeling approach, outlined in Chapter 1, closely follows the discrete choice literature with an emphasis on allowing tastes to vary across locations. The importance of flexibly modeling heterogeneity in discrete choice setups has been well documented in the literature (see Berry, Levinsohn, and Pakes (1995), Petrin (2002), Song (2007)). The model allows for the fact that, for example, the removal of a popular sandal will be much more costly for markets in Florida than for markets in Alaska. Failing to account for this heterogeneity will place heavy dependence on the idiosyncratic logit error, resulting in estimated welfare benefits of variety that are much too large.

Employing our data at the level of narrowly defined products and at narrow geographic detail, however, also presents us with an empirical challenge. Despite the fact that we observe millions of sales, the large number of products and locations, inevitably leads to many products having local market shares equal to zero. For example, even in the annualized data 82.2% of products have zero sales at the metro level and 59.8% of products have zero sales at the state level.[2] As highlighted in Chapter 1, these zeros are problematic for two reasons. First, using standard estimation techniques create a selection bias in the demand estimates, leading to incorrect estimates of the gains to variety. Second, the zeros suggests a small samples problem that, if uncorrected, would lead us to overstate the degree of heterogeneity across markets and understate the gains to variety.[3]

While we could further aggregate over either geography or product space to reduce the percentage of zeros, this would be unsatisfactory because it would significantly obscure the across-market heterogeneity of interest to us. Instead we address this problem using the estimation technique proposed in Chapter 1. Our estimation strategy exploits the structure of the demand model to, in a sense, separate the problem into two parts. Using the local market share information to form a set of micro-moments that augment

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

[3] For example, suppose on a particular day, a local market sees a single shoe sale and that is the only sale of that particular shoe nationwide. If we fail to account for the small samples issue, we would come to the conclusion that 1) the rest of the country has absolutely no interest in the product, and 2) that local market would not benefit from any additional variety because consumers only want that particular shoe.

the aggregated (national) sales data, which allows us to retain valuable information on across-market demand heterogeneity while explicitly accounting for small samples. Then at the aggregate level, our approach effectively mimics the standard approach and we are able to pin down the price coefficient and other parameters common across markets.

Since local choice sets are often unobserved, there is the additional challenge of forecasting local choice sets for counterfactual analysis. As mentioned above, brick-and-mortar retailers tend to cater their assortments to local demand. Using our estimated demand, we infer which products local brick-and-mortar retailers would be stocking in the absence of online retail. Unfortunately, because of the number of products, the combinatorial problem of choosing the most profitable assortment of items becomes intractable. Consistent with the literature, we will assume local brick-and-mortar retailers stock the top K most popular products. This is determined by the estimated local mean utilities from the demand system. We can then calculate the consumer welfare in worlds with and without the online retailer.

Our results indicate that demand for specific products varies significantly across markets and we show that accounting for this heterogeneity is necessary for rationalizing the distribution of local sales. When brick-and-mortar retailers cater their assortments to local demand, we find that the welfare gains from online variety are relatively small. From 6.4 to 9.4% of the total unconstrained consumer welfare is due to online variety. However, if we shut down the across-market demand heterogeneity, and hence the localization in brick-and-mortar retail, we would find 16.3 to 31.7% of the unconstrained consumer welfare is due to online variety, an overstatement of between 187 and 338%. Put another way, if local stores cater to local demand, then the value of online markets is relatively small because the average consumer already has access to the products they want to purchase. Additionally, for brick-and-mortar retailers, we find a large incentive for them to cater to their local demand. By doing so they can obtain 14.5 to 28.8% higher revenue than under a standardized assortment.

Finally, our results suggest a new interpretation of the "long tail" phenomenon observed in online retail (Anderson 2004). The term describes a shift in the distribution of revenue toward niche, or tail, products. The prevailing view is that the long tail pattern has emerged because niche products better satisfy consumer tastes, leading consumers to switch from purchasing hit products available at their local brick-and-mortar retailers to purchasing niche products only available online.[4] [5] This suggests sizable welfare gains by increasing variety. However, we find relatively smaller gains from increasing the variety of shoes. This is due to the fact that in the presence of across-market demand heterogeneity, aggregating demand across markets yields a long tail nationally, but at the local level, the tails are relatively short which traditional retailers can capture through localization of assortments.[6]

The rest of the chapter will be organized as follows. Section 2 discusses our data and presents preliminary evidence of across-market heterogeneity. In Section 3, we present the model. Section 4 discusses our estimation procedure to be followed by our demand results in Section 5. Section 6, contains our counterfactual analysis. Section 7 discusses the robustness of our findings and Section 8 concludes the paper.

2.2 Data

We create several original data sets for this study. The main data set consists of detailed point-of-sale, product review, and inventory data that we collected from a large online retailer. With this data, we observe over $1 billion worth of online shoe transactions between 2012 and 2013. We augment this with a snapshot of shoe availability for two brick-and-mortar retailers, Macy's and Payless ShoeSource. A discussion of this data can be found in Appendix B.1.

We begin by summarizing our data (Section 2.2.1), then we provide evidence of

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

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

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

across-market consumer demand heterogeneity (Section 2.2.2). Finally, we document the "zeros problem" in the data and discuss aggregation as a means to address the issue (Section 2.2.3).

2.2.1 Online Shoe Sales

The main data for this study was collected and compiled with permission from a large online retailer. This online retailer sells a wide variety of product categories, including footwear, which will be the focus of our analysis. Each transaction in the point-of-sale (POS) data base contains the timestamp of the sale, the 5-digit shipping zip code, price paid, and a wealth of information about the shoe. Each sale corresponds to a stock-keeping unit (SKU) and a numeric code for the style. The style code allows us to discern red versus blue of the same shoe model. The transaction identifier allows us to see if a customer purchased more than a single pair of shoes. For each product we record the brand, product material, and many categorical classifying variables, such as if a shoe is a wingtip and the material of the shoe. Finally, we download a picture of each shoe, and image process them to create color covariates.

We also merge in product review and inventory data. The review data contains the time series of reviews for each SKU. Each review contains reported ratings on comfort, look, and overall appeal. For the inventory data, we track daily inventory for every shoe.[7] 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.

We observe over 13.5 million shoe transactions during the collection period, with a majority of transactions being women's shoes. The price of shoes varies substantially across gender, but also within gender – for example, dress shoes tend to be more expensive than walking shoes. The distribution of transaction size per order is heavily skewed to the left. Only a very 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

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

discrete choice problem.

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

An important feature of the data is the number of products the online retailer offers. The average daily assortment size is over 50,000 products, and over the span of data collection, over 100,000 varieties of shoes were offered for sale. This constantly changing choice set provides us with additional variation that will help us identify the parameters of our model.

2.2.2 Across-Market Demand Heterogeneity

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

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

Since prices, product characteristics, and choice sets are the same for all markets, differences in observed local market shares can only be rationalized by differences in local demand. In Table 2.1, we present the average local and national share of revenue generated by the top 1,000 products ranked by local market. If demand was homogeneous across markets, we would expect the share of revenue accruing to these products to be the same locally and nationally. Thus, the two columns of Table 2.1 would be equal. Instead we see the share of revenue generated by these products are very large at the local market level compared their share of revenue at the national level. For example, the top 1,000 products ranked at the metro (combined statistical area - CSA) level make up 86.9% of revenue at the metro level, but these same products only accounts for 11.5% of national revenue. This suggests that the commonality, even among the most popular products, is quite small across markets.

Table 2.1: Revenue Share of Top 1,000 Products

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

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

We can formally test for across-market demand heterogeneity using multinomial tests comparing local market shares ($s_{\ell j}$) to national market shares (s_j), where the null hypothesis is $H_0 : s_{\ell j} = s_j$, for all $j \in J$. Table 2.2 presents the rejection rates for various levels of aggregation. We can see that these tests are overwhelmingly rejected at all levels of aggregation. However, in the tests at the monthly level, we can see the effects of both zeros and aggregation beginning to appear. At more disaggregated levels, zeros become more prevalent, reducing the power of the multinomial tests. On the other end of the spectrum, aggregating up to Census Regions greatly obscures heterogeneity across markets leading to a reduction in rejection rates when compared to the Census Division level.

Some differences across markets occur for obvious reasons. Take our earlier example of boots versus sandals. Figure 2.1 plots the predicted values from a regression of a state's average annual temperature on the share of state revenue captured by boots (dashed) and sandals (solid). As expected, boots make up a greater share of revenue in colder states and a smaller share in warmer states. Conversely, the opposite relationship

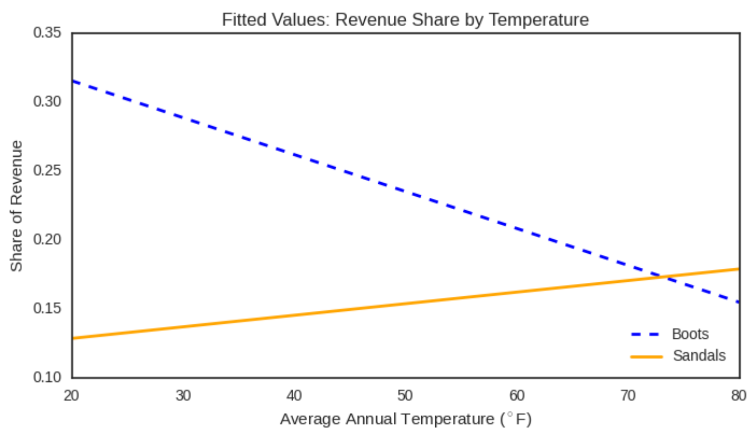
Table 2.2: Multinomial Tests - Rejection Rates

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

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

holds for sandals.[8]

Figure 2.1: Boots vs. Sandals Revenue by Temperature



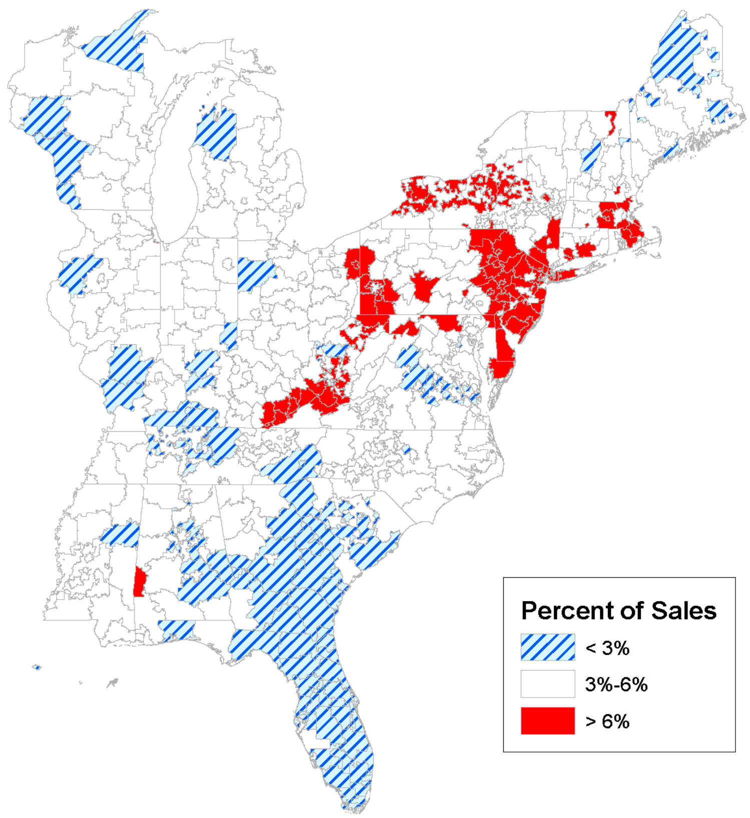
Other differences across markets occur for less obvious reasons. In Figure 2.2, we map the consumption pattern of a popular brand, measured by national revenue. Annual sales are mapped into 3 digit zip codes for the eastern United States.[9] While this brand tends to be popular over a large portion of the country, we can see a clear preference

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

[9] We isolate the eastern United States to be able to distinguish differences at fine levels of disaggregation and because the interesting portion of the map happens to be the northeastern part of the country. The full map can be viewed in Appendix B.3 (Figure B.4).

for this brand in the northeast. In Florida this brand makes up less than 2.5% of sales, while in parts of New York, New Jersey, and Massachusetts it makes up over 6% of sales. We will exploit this variation to help us identity across-market demand heterogeneity.

Figure 2.2: Sales Share of a Popular Brand Across Zip3s



2.2.3 Aggregation and the Zeros Problem

While demand varies across locations, the data at disaggregated levels exhibits a severe small samples problem, which manifests itself in the form of a zeros problem. Table 2.3 illustrates the effect of disaggregating the data across both geography and time. For each product, an observation is the number of sales by geographic area and time horizon. We then calculate the percentage of observations where no sale is observed. For example, at the CSA level 95% of products have zero monthly sales. This highlights a small samples problem that is common in high frequency sales data and data with large choice sets.

Observations of zero sales is problematic from both a theoretical and empirical point of view. The distributional assumptions in many demand models imply no products will have zero sales and empirical techniques often require taking the natural log of sales or market shares. This, of course, does not exist when zero sales are observed and common solutions to this problem will lead to biased demand estimates. An in-depth discussion of these issues can be found in Berry, Linton, and Pakes (2004), Gandhi, Lu, and Shi (2013), and Gandhi, Lu, and Shi (2014).

Table 2.3: Data Disaggregation: The Zeros Problem

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

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

On the other hand, aggregation can resolve some of the small samples issue, but it is unsatisfactory because it significantly obscures across-market heterogeneity. For example, we could further aggregate over geography to the Census Region, which would reduce the percentage of zeros to 23.3%. However, this would also reduce the number of markets to four and yet, the percentage of zeros is still quite high, and further aggregation would be necessary. Furthermore, from Table 2.1, we can see that the top 1,000 products ranked at the Census Region level make up a similar share of revenue at the market and the national levels. That is, we have greatly obscured the across-market heterogeneity of interest to us.

We could also aggregate over product space. Table 2.4 shows the percentage of zeros and the local and national revenue shares of the top products ranked by local market for products at the SKU-style level (our definition of a product) and aggregated to the SKU, brand-category, and brand levels. Since aggregating to the brand-category and brand levels greatly reduces the number of products, we adjust the benchmark to the top 10 "products," rather than the top 1,000.

Table 2.4: Revenue Share of Top Products: Product Aggregation

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

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

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

2.3 Model

This section presents our demand models for online shoes. We consider the simple and nested logit specifications introduced in Chapter 1. A well known criticism of the simple logit model is that it assumes consumer taste shocks are independent across the available choices. In particular, it precludes consumers from having correlated preferences for categories of shoes, such as boots over sandals. The advantage of the nested logit model is that it allows these types of preferences by allowing consumer tastes to be correlated within category nests.

The choice set of available shoes is the same across locations, which we will define to be CSA's, but varies by time period, as some products enter and others exit each period. Suppressing the time subscript, let $J \cup \{0\}$ be the set of available shoes with $\{0\}$ denoting the outside good. Shoes can also be categorized into mutually exclusive and exhaustive sets $c \in C \cup \{0\}$, where the outside good is contained in its own nest. Categories of men's shoes include boat, boots, clogs, oxfords, sandals, slippers, and sneakers. Women's shoes have two additional categories, flats and heels.

Each consumer solves a discrete choice utility maximization problem: Consumer $i \in I_\ell$ in location $\ell \in L$ will purchase a shoe j if and only if the utility derived from shoe j is greater than the utility derived from any other shoe, $u_{i\ell j} \geq u_{i\ell j'}, \forall j' \in J \cup \{0\}$. For a shoe $j \in J \cup \{0\}$, the utility of a consumer $i \in I$ in location $\ell \in L$ is given by

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

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

$$\nu_{i\ell j} = \begin{cases} \eta_{\ell j} + \varepsilon_{i\ell j} & \text{(Simple)} \\ \eta_{\ell c} + \eta_{\ell j} + \zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j} & \text{(Nested)} \end{cases},$$

where $\eta_{\ell j}$ is drawn independently from a normal distribution, $N(0, \sigma_j^2)$, $\eta_{\ell c}$ is a location-category fixed effect, $\varepsilon_{i\ell j}$ is drawn i.i.d. from a Type-1 extreme value distribution, and ζ_{ic} has the unique distribution such that $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$ has a Type-1 extreme value distribution. For a consumer i, ζ_{ic} is common to all shoes in the same category and has a distribution that depends on the nesting parameter λ, $0 \leq \lambda < 1$. λ determines the within-category correlation of utilities. These terms decompose the heterogeneity in the random utility among consumers into an "across-market" effect, $\eta_{\ell j}$ or $\eta_{\ell c} + \eta_{\ell j}$, and a "within-market" effect, $\varepsilon_{i\ell j}$ or $[\zeta_{ic} + (1-\lambda)\varepsilon_{i\ell j}]$. When $\eta_{\ell c} = \eta_{\ell j} = 0$ for all $\ell \in L, c \in C, j \in J$, then the model reduces to a standard "love of variety" simple or nested logit model, where there is no distinction between local and national preferences.

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

$$\pi_{\ell j} = \pi_j(\eta_\ell;\delta) = \begin{cases} \frac{\exp\{\delta_j+\eta_{\ell j}\}}{\sum_{j'=0}^{J}\exp\{\delta_{j'}+\eta_{\ell j'}\}} & \text{(Simple)} \\ \frac{\left(\sum_{j\in c}\exp\{\delta_{\ell j}/(1-\lambda)\}\right)^{1-\lambda}}{1+\sum_{c'\in C}\left(\sum_{j'\in c'}\exp\{\delta_{\ell j'}/(1-\lambda)\}\right)^{1-\lambda}} \cdot \frac{\exp\{\delta_{\ell j}/(1-\lambda)\}}{\sum_{j'\in c}\exp\{\delta_{\ell j'}/(1-\lambda)\}} & \text{(Nested)} \end{cases}. \tag{2.1}$$

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

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

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

2.4 Estimation

Since we have a large number of locations, 165 CSA's, for any $\sigma = \{\sigma_j\}_{j=1}^{J}$, we can exploit the structure of the model by simulating $\tilde{\eta}_{\ell j} \sim N(0,\sigma_j^2)$ to approximate,

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

With aggregated choice probabilities depending only on the variance of the across-market heterogeneity, σ, national shoe demand can be expressed as

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

Propositions 1 and 2 in Section 1.3.2, show how to invert this system of equations to yield,

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

where x_j is a vector of shoe characteristics, p_j is the price of shoe j, and ξ_j is the unobserved product quality for shoe j. Included in x are ratings for comfort, look, and overall appeal, as well as, fixed effects for color, brand, and time.

From Propositions 1 and 2, the market share inversions take the following form

$$\delta_j = \begin{cases} \log\pi_j - \log\sum_{\ell=1}^{L}\omega_\ell\pi_{\ell 0}\exp\{\tilde{\eta}_{\ell j}\} & \text{(Simple)} \\ (1-\lambda)\left(\log(\pi_j) - \log\left(\sum_{\ell\in L}\omega_\ell\pi_{\ell c}\left(\frac{\pi_{\ell 0}}{\pi_{\ell c}}\right)^{\frac{1}{1-\lambda}}\exp\left\{\frac{\tilde{\eta}_{\ell j}}{1-\lambda}\right\}\right)\right) & \text{(Nested)} \end{cases},$$

where the shoe's national market share, π_j, location-specific population shares, ω_ℓ, location-specific shares of the outside good, $\pi_{\ell 0}$, and location-specific shares of each shoe category, $\pi_{\ell c}$, are known from the data. Then, for a fixed σ and λ, we can use linear instrumental variables z_j, such that $E[z_j \xi_j] = 0$ and $E[z_j'(p_j, x_j)]$ has full rank, to identify (α, β) as a function of σ and λ.

We instrument for price using the characteristics of competing products (BLP instruments), grouped by brand. That is, let B denote the set of brands and let J_b denote the set of shoes manufactured by brand $b \in B$, then, for each time period, our set of price instruments is

$$x_j, \quad \sum_{j' \neq j \in J_b} x_{j'}, \quad \sum_{j' \in J_{-b}} x_{j'}.$$

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

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

To identify the extent of across-market demand heterogeneity, σ, we employ the micro-moments proposed in Section 1.3.1. Define $P0_{\ell j}(\sigma)$ to be the probability that a shoe j has zero sales given the N_ℓ consumers observed to purchase a shoe in location ℓ. We then define,

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

to be the fraction of markets that the model predicts will have zero sales for shoe j. 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 shares for shoe j. Our micro moment then identifies σ by matching the model's prediction to the empirical analogue, i.e.

$$m(\sigma) = \sum_{j=1}^{J} \left(P0_j(\sigma) - \hat{P}0_j \right)^2.$$

To ease the estimation, we parameterize σ in the following way

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

where σ_j depends on shoe j's category. All demand parameters are estimated jointly using GMM. A further description of the estimation procedure and identification arguments are found in Sections 1.3.3 and 1.3.4, respectively.

2.5 Results

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

2.5.1 Demand Parameters Constant Across Markets

A summary of our demand estimates is presented in Tables 2.5 and 2.6 for men's and women's shoes, respectively. Each specification includes fixed effects for brand, color, and time. We present six sets of estimates: (1) the simple logit demand model estimated at the CSA level, which we will call "local simple logit," (2) the simple logit demand model estimated at the national level, which we will call "national simple logit," and (3) our estimation procedure, which we will call "QW simple logit." Specifications (4), (5), and (6) are the equivalent for the nested logit setup. We also account for any remaining zeros using the correction proposed by Gandhi, Lu, and Shi (2014). A discussion of the correction procedure can be found in Appendix B.2.

Specifications (1) and (4), the simple and nested logit demand models estimated at the local level, illustrate the selection bias generated by the severity of the zeros problem. When estimating the models at the CSA level, each observation is a product-location specific share. Thus, the number of observations in the local simple and nested

Table 2.5: Demand Estimates - Men's

	Simple Logit			Nested Logit		
	Local	National	QW	Local	National	QW
	(1)	(2)	(3)	(4)	(5)	(6)
Price	-0.011	-0.019	-0.034	-0.029	-0.019	-0.024
	(0.000)	(0.001)	(0.003)	(0.001)	(0.000)	(0.002)
Comfort	0.030	0.102	0.085	0.004	-0.018	-0.012
	(0.004)	(0.017)	(0.027)	(0.005)	(0.011)	(0.015)
Look	-0.112	-0.171	-0.175	-0.136	-0.062	-0.080
	(0.004)	(0.017)	(0.031)	(0.005)	(0.012)	(0.018)
Overall	0.214	0.436	0.530	0.209	0.249	0.312
	(0.004)	(0.018)	(0.055)	(0.006)	(0.012)	(0.031)
No Reviews	0.289	0.769	1.485	0.356	0.713	0.936
	(0.013)	(0.067)	(0.160)	(0.018)	(0.044)	(0.091)
Constant	-13.472	-16.745	-17.406	-10.924	-10.091	-11.419
	(0.030)	(0.119)	(0.519)	(0.081)	(0.142)	(0.299)
λ	—	—	—	0.129	0.602	0.512
				(0.004)	(0.011)	(0.006)
σ	—	—	$*$	—	—	$*$
Fixed Effects						
Brand	✓	✓	✓	✓	✓	✓
Color	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓
N	28,916,910	175,254	175,254	28,916,910	175,254	175,254
Zeroes	27,597,784	30,282	30,282	27,597,784	30,282	30,282
	(95.4%)	(17.3%)	(17.3%)	(95.4%)	(17.3%)	(17.3%)
Price Elast.						
Mean	-1.035	-2.206	-3.828	-3.054	-5.501	-4.833
Std. Dev.	(0.591)	(1.637)	(2.841)	(1.745)	(4.082)	(4.033)

Notes: Estimated at the monthly level. "Local Simple Logit" (1) and "Local Nested Logit" (4) estimate the simple and nested logit models at the CSA local market level, hence the ξ's are market level fixed effects. "National Simple Logit" (2) and "National Nested Logit" (5) estimate the simple and nested logit models at the national level. Finally, "QW Simple Logit" (3) and "QW Nested Logit" (6) estimate the simple and nested logit models using our estimation technique to allow for across-market heterogeneity.

Standard errors in parentheses. All reported coefficients are significant at the 1% level.

$*$ estimates for across-market heterogeneity in specifications (3) and (6) will be discussed in the following subsection.

Table 2.6: Demand Estimates - Women's

	Simple Logit			Nested Logit		
	Local	National	QW	Local	National	QW
	(1)	(2)	(3)	(4)	(5)	(6)
Price	0.001	-0.008	-0.011	0.001	-0.003	-0.005
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Comfort	0.105	0.053	0.084	0.108	-0.021	0.008
	(0.002)	(0.011)	(0.025)	(0.003)	(0.004)	(0.001)
Look	-0.052	-0.096	-0.309	-0.054	-0.014	-0.044
	(0.003)	(0.011)	(0.025)	(0.003)	(0.004)	(0.013)
Overall	0.087	0.368	0.430	0.094	0.025	0.073
	(0.003)	(0.013)	(0.028)	(0.003)	(0.006)	(0.004)
No Reviews	0.136	0.426	0.323	0.147	0.035	0.076
	(0.008)	(0.039)	(0.008)	(0.009)	(0.014)	(0.001)
Constant	-14.450	-17.938	-19.994	-14.629	-17.938	-11.614
	(0.011)	(0.052)	(0.213)	(0.014)	(0.052)	(0.097)
λ	—	—	—	-0.026	0.863	0.648
				(0.001)	(0.008)	(0.024)
σ	—	—	$*$	—	—	$*$
Fixed Effects						
Brand	✓	✓	✓	✓	✓	✓
Color	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓
N	57,682,020	349,588	349,588	57,682,020	349,588	349,588
Zeroes	55,141,243	66,827	66,827	55,141,243	66,827	66,827
	(95.6%)	(19.1%)	(19.1%)	(95.6%)	(19.1%)	(19.1%)
Price Elast.						
Mean	0.118	-0.949	-1.262	0.108	-2.866	-1.525
Std. Dev.	(0.077)	(0.813)	(1.081)	(0.071)	(2.456)	(1.339)

Notes: Estimated at the monthly level. "Local Simple Logit" (1) and "Local Nested Logit" (4) estimate the simple and nested logit models at the CSA local market level, hence the ξ's are market level fixed effects. "National Simple Logit" (2) and "National Nested Logit" (5) estimate the simple and nested logit models at the national level. Finally, "QW Simple Logit" (3) and "QW Nested Logit" (6) estimate the simple and nested logit models using our estimation technique to allow for across-market heterogeneity.
Standard errors in parentheses. All reported coefficients are significant at the 1% level.
$*$ estimates for across-market heterogeneity in specifications (3) and (6) will be discussed in the following subsection.

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

Specifications (2) and (3) directly compare the results estimated using the standard approach on national level market shares and the results estimated using our procedure for the simple logit model. Similarly, specifications (5) and (6) compare the estimates for the nested logit model. Unsurprisingly, the results for these approaches are similar. However, the advantage of our approach is that it retains information on the distribution of heterogeneity across locations. The importance of this distinction will be highlighted in the following section when we perform counterfactual analyses at the local level.

In the simple logit model estimated using our estimation technique (QW), the price coefficients have the expected signs, -0.034 and -0.011 for men's and women's shoes, respectively. These results suggest that men are far more price sensitive than women when it comes to their footwear purchases with mean product level price elasticities of -3.828 versus -1.262. While price coefficients in the nested logit specification are smaller in magnitude when compared to the simple logit price coefficients, -0.024 and -0.005 for men's and women's shoes, respectively, the additional flexibility of the nested logit model has the expected effect on consumer substitution patterns. By allowing consumer level taste shocks within categories of shoes to be correlated, the nested logit model reduces the dependence on the idiosyncratic logit error. The result is price elasticities that are larger in magnitude, -4.833 and -1.525 for men's and women's shoes, respectively.

Turning to the coefficients on our review variables, we can see that the overall rating has the expected sign in all specifications, with higher ratings having positive effects on demand. Look and comfort, however, appear to have a sign that is opposite from the one expected in at least some specifications. It is possible that, after controlling for the overall rating, the qualities that make a shoe more aesthetically pleasing or more comfortable reduces a shoe's appeal through some other channel. Meanwhile, our indicator for no reviews takes on positive signs for both men's and women's shoes. This variable largely captures the demand for new products before there has been an opportunity to review them. New products often benefit from additional promotion and advertising and it is likely that the positive effect of having no review actually reflects additional promotion, rather than a desire to purchase shoes that have not been reviewed.

2.5.2 Across-Market Heterogeneity

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

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

by matching the predicted (by the model) percentage of locations in which a product has zero sales with the observed percentage of locations that have zero sales. Our estimates for the across-market demand heterogeneity are presented in Table 2.7. For comparison, the interquartile range and the standard deviation of mean utilities, δ, are also reported.

In general, women's shoes tend to exhibit greater across-market demand heterogeneity than men's shoes. This is due, in part, to the greater number of women's shoe varieties. Additionally, the the estimated across-market demand heterogeneity tends to be smaller for the nested logit model. This is because the σ in the nested logit setup is conditional on the differences in demand for categories across locations, whereas in the simple logit setup differences in category demand across locations get soaked up by σ. For example, suppose boat shoes had a 90% market share in location 1 and a 10% market share in location 2. The simple logit setup would require a large σ to rationalize these differences. However, conditional on the overall demand for boat shoes being smaller in location 2, the differences in demand for individual products, within that category, across these location may be small.

Under the simple logit model, the men's categories with the largest σ's are slippers (2.371) and sneakers (1.944), while the women's categories with the largest σ's are boat (5.017), slippers (3.376), and sneakers (4.065). In the nested logit model, the categories with the largest across-market heterogeneity, conditional on category choice,

Table 2.7: Results: Across-Market Heterogeneity: $\sigma_j = h(\cdot)$

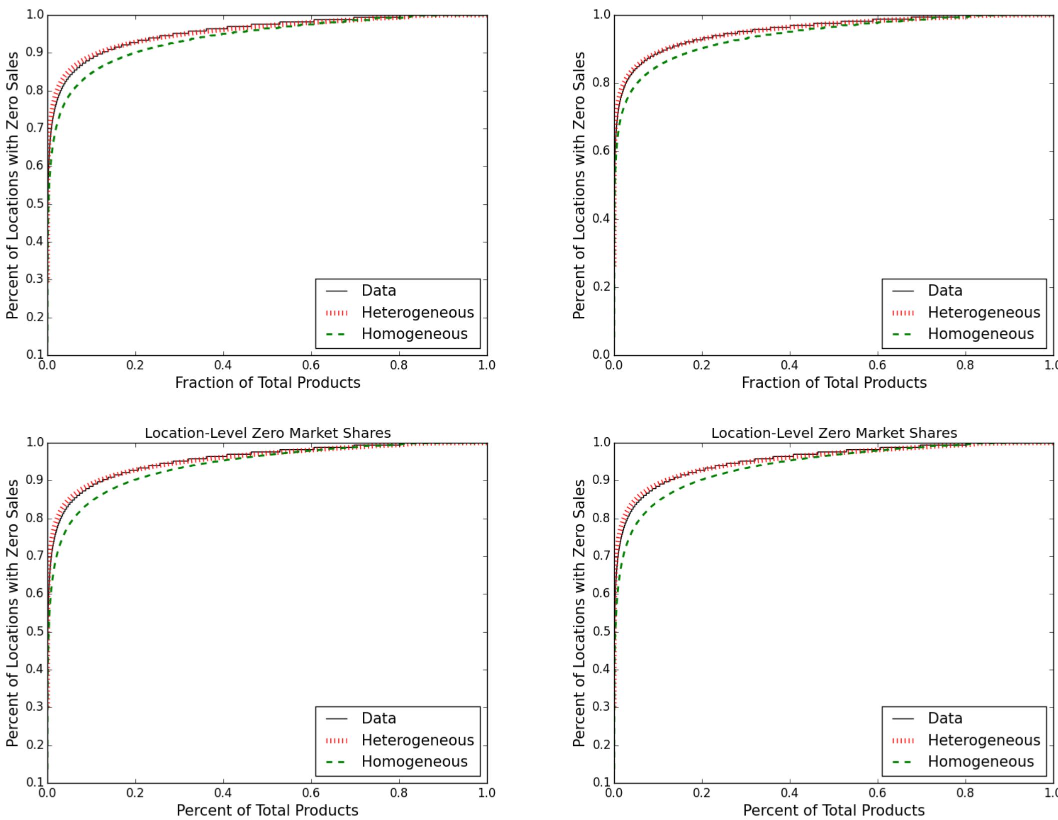
	Simple Logit		Nested Logit	
	Men	Women	Men	Women
Boat	1.311	5.017	0.652	0.879
Boots	0.505	1.777	0.248	0.798
Clogs	0.007	1.524	0.246	0.829
Flats	—	0.536	—	0.582
Heels	—	0.644	—	0.670
Oxfords	0.501	1.798	0.822	0.262
Sandals	0.348	0.305	0.568	0.190
Slippers	2.371	3.376	0.952	1.285
Sneakers	1.944	4.065	0.815	0.213
δ				
IQR	2.988	4.835	1.720	2.320
St. Dev.	3.050	4.089	1.628	1.901

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

are oxfords (0.822), slippers (0.952), and sneakers (0.815) for men and slippers (1.285) for women. These appear to be quite large with across-market variation in some categories approaching the magnitude of the variation in the aggregate mean utilities of the simple logit model and approaching about half the magnitude of the nested logit model.

Figure 2.3 illustrates the fit of our estimation technique and gives us further insight into our heterogeneity results. It plots the percentage of location level zero market shares by product. The left panels are plots for men's shoes and the right panels are for women's shoes. The top panels are for simple logit and the bottom panels are for nested logit. For comparison, we include simulation results for the case of homogeneous demand across markets, i.e. when $\sigma_j = 0$ for all shoes j. At the head of the distribution there are fewer location level zero market shares, but, because mean utilities are relatively high, variation is required to produce these zeros. Moving toward the middle of the distribution, this variation is able to account for the increasing percentage of zero market shares. If demand were homogeneous across markets, we would expect to see far fewer zeros among popular and mid-ranked products.

Figure 2.3: Goodness of Fit: Percentage of Location Level Zeros



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

2.6 Analysis of the Estimated Model

In this section, we use the estimated model to perform counterfactual analysis under a series of restricted choice sets. We will begin by presenting our primary results, allowing for tastes to differ across markets and for local brick-and-mortar retailers to customize their assortments to local demand (Section 2.6.1). We will then present results shutting down local assortment customization and show how these results overestimate the gains to online variety (Section 2.6.2). Finally, we revisit the phenomenon of the long tail and show that aggregation of sales over markets with different tastes is a key driver of the long tail of online retail (Section 2.6.3).

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

Mechanically, to compute our counterfactuals, we draw a set of $\eta_{\ell j}$'s for each location. Products are then ranked in each location by their location specific mean utilities and the top products are included in the counterfactual choice set. For each counterfactual choice set, location level choice probabilities are then calculated according to Equation 2.1. Using these probabilities, we simulate location level purchases, which then allow us to compute counterfactual consumer welfare and retail revenue.

2.6.1 Counterfactuals with Across-Market Heterogeniety

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

$$CS_\ell = \begin{cases} \frac{M\omega_\ell}{\alpha} \log\left(1 + \sum_{j=1}^{J} \exp\{\delta_j + \eta_{\ell j}\}\right) & \text{(Simple)} \\ \frac{M\omega_\ell}{\alpha} \log\left(1 + \sum_{c \in C} \left(\sum_{j \in c} \exp\{\delta_j + \eta_{\ell j}\}\right)^{1-\lambda}\right) & \text{(Nested)} \end{cases},$$

and retail revenue,

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

where M is the size of the national population. For each of our specifications, Table 2.8 presents the increase in consumer welfare from online variety and Table 2.9 presents the consumer welfare and retail revenue under the restricted choice set relative to the unconstrained online choice set.

Examining the results of our estimation technique, we find that when consumers are able to move from a world where they only have access to the goods available at their local stores to a world where they have access to the whole online choice set, consumer welfare increases by 10.4% or $57.69 million under the simple logit model and by 6.8% or $57.20 million under the nested logit model.

Table 2.8: Local Choice Set: Consumer Welfare Increase

	Simple Logit		Nested Logit	
	Local	QW	Local	QW
Increase				
- Absolute ($ Millions)	0.02	57.69	0.72	57.20
- Percent	$\sim$ 0%	10.4%	$\sim$ 0%	6.8%

The deficiencies of using the standard approach with location level market shares is also highlighted by Table 2.8. Consistent with Ellison and Glaeser (1997), ignoring the local level small samples problem and employing a local level simple or nested logit estimation exaggerates the extent of heterogeneity across markets. By assuming products without an observed sale are completely unwanted at that particular location, it is "easier" for our hypothetical brick-and-mortar retailers to customize their assortments to local demand. This leads us to understate the gains to consumer welfare. In our application, using the standard approach with location level market shares implies that the consumer welfare gains from online variety are almost nonexistent. This is because, in our data, relatively few shoe varieties make up the observed sales in individual local markets. Thus relatively few shoe varieties are required to completely satisfy local demand when ignoring the small samples problem. Note that we omit the national level specifications. While these specifications may be consistent with across-market demand heterogeneity, there is no way to determine the underlying geographic distribution of heterogeneity.

Table 2.9: Local Choice Set: Share of Unconstrained

	Simple Logit		Nested Logit	
	Local	QW	Local	QW
Consumer Welfare	$\sim$ 100%	90.6%	$\sim$ 100%	93.6%
Revenue	$\sim$ 100%	89.8%	$\sim$ 100%	92.2%

Table 2.9 suggests that consumer welfare derived from access to online variety may be surprisingly small. Using our preferred specifications, QW simple and nested logit, if local stores stock products that target local demand, consumers would capture 90.6% to 93.6% of the unconstrained consumer welfare - the total consumer welfare that would be obtained with access to all of the products. Conversely, having access to the entire online choice set only accounts for 6.4% to 9.4% of the total unconstrained consumer welfare. Similar conclusions can be drawn for retailer revenue. A national brick-and-mortar chain can generate 89.8% to 92.2% of the total revenue it would generate by stocking the universe of products, by stocking a only small number of well selected products. Also, Table 2.9 again highlights the effects of ignoring the small samples problem. Similar to the result for consumers, estimates ignoring the small samples problem imply almost zero benefits to firms of stocking the entire universe of varieties.

2.6.2 Counterfactuals with Nationally Standardized Choice Sets

In this subsection, we perform counterfactual analyses similar to the ones above. However, we impose the additional constraint that each market will be restricted to the top products determined by ranking products according to their national mean utilities, δ_j. Using our preferred estimates, Table 2.10 presents the increase in consumer welfare from online variety and Table 2.11 presents the consumer welfare and retail revenue under the restricted choice set relative to the unconstrained online choice set.

Table 2.10: National Choice Set: Consumer Welfare Increase (QW Estimates)

	Local Choice Set		National Choice Set	
	Simple	Nested	Simple	Nested
Increase				
- Absolute ($ Millions)	57.69	57.20	194.68	146.63
- Percent	10.4%	6.8%	45.5%	19.5%

Table 2.10 shows that failing to account for customization in local assortments will overstate the gains to consumer welfare. Under a standardized national assortment, access to online variety increases consumer welfare by 45.5% or $194.68 million in the simple logit model and by 19.5% or $146.63 million in the nested logit model. This suggests failing to account for heterogeneity across markets will overstate consumer welfare due to online variety by 338% in percentage terms and 237% in absolute terms under the simple logit model and by 187% in percentage terms and 156% in absolute terms under the nested logit model. The overstatement occurs because the initial welfare (pre-internet) of consumers is lower when choice sets are nationally standardized than when they are locally targeted, which can be seen in Table 2.11. That is, failing to allow for local customization in assortments will overstate how bad the world was before the internet.

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

Table 2.11: National Choice Set: Share of Unconstrained (QW Estimates)

	Local Choice Set		National Choice Set	
	Simple	Nested	Simple	Nested
Consumer Welfare	90.6%	93.6%	68.3%	83.7%
Revenue	89.8%	92.2%	69.7%	80.5%

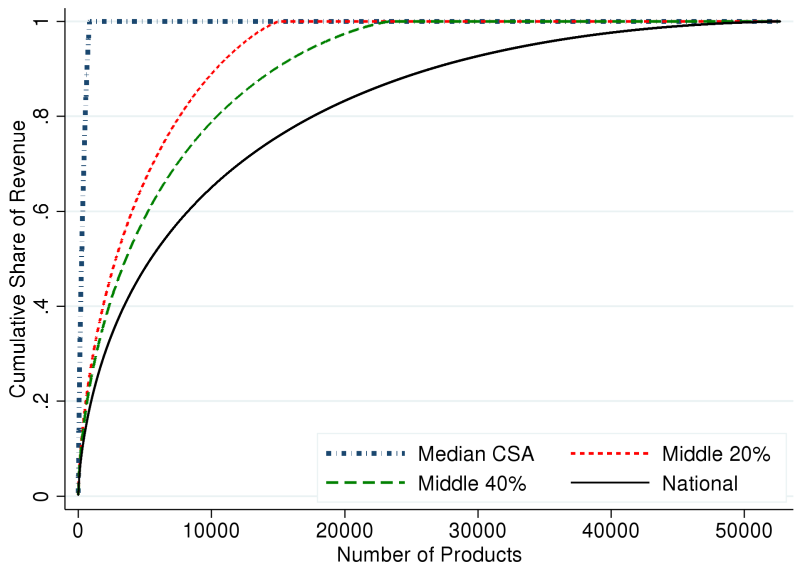
Turning to retailer revenue we see that a researcher assuming nationally standardized assortments will severely underestimate counterfactual brick-and-mortar revenue. Under the simple logit specification, a national brick-and-mortar chain would generate just 69.7% of total unconstrained revenues with a nationally standardized assortment compared to 89.8% of total unconstrained revenues with locally customized assortments. For nested logit, these numbers are 80.5% and 92.2%, respectively. This suggests that there is a significant incentive for local stores to cater to local demand. By doing so they would obtain 14.5% to 28.8% greater revenue than under a nationally standardized assortment.

2.6.3 Long Tail Analysis

Our counterfactual results in the previous two subsections suggest that "shorter" tails at the local level underly the long tail at the national level. Using the raw sales data, Figure 2.4 illustrates how local level "short" tails can aggregate to a national level long tail. It plots the cumulative share of revenue going to the top K products for the median CSA (by number of monthly sales), middle 20%, middle 40%, and national level markets. For a single local market, we can see that there is an extremely short tail with fewer than 3,000 products making up all the of sales in that CSA. Since the popularity of products varies wildly across geographic markets, aggregating over markets increases the number of different varieties sold and decreases the density of sales among the top ranked products. Sales becoming less concentrated among the top products produces a lengthening effect on the tail of the sales distribution.

However, the small samples problem in the raw data presents us with a skewed

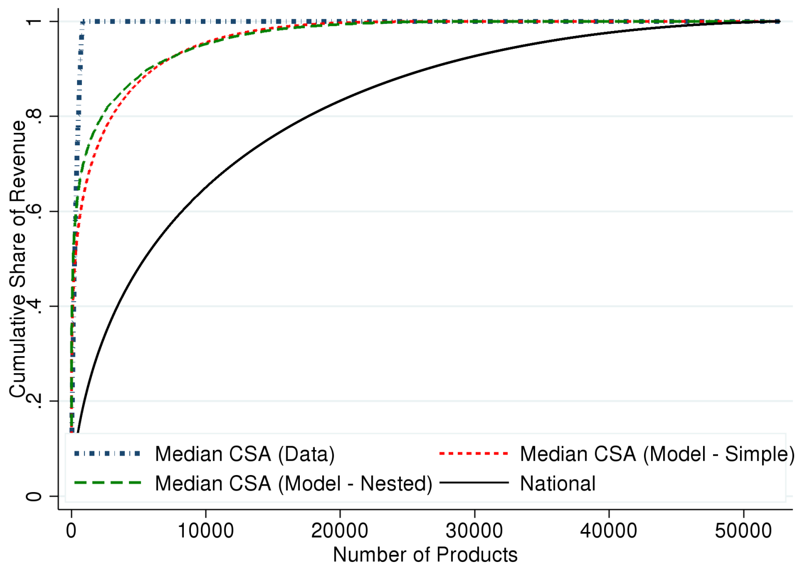
Figure 2.4: Aggregating to the Long Tail



perspective in that it suggests a ridiculously short tail at the local level. Using our estimated model, we can correct for the small samples problem in our long tail analysis by simulating a large number of sales in these markets. Figure 2.5 plots the cumulative share of revenue going to the top K products for the median CSA correcting for the small samples problem. As expected, we find that the local tail is actually quite a bit longer than suggested by the raw data.

Table 2.12 further illustrates the effect of small samples on the local tail. It presents the average share of revenue accruing to products outside of the top 3,000 products. At the national level, more than 50% of revenue comes from products ranked outside the top 3,000. At the local level, if we were to rely on the raw data, we would find that only 3.4% of revenue comes from products ranked outside the top 3,000 products. In other words, 96.6% of demand could be satisfied with just 3,000 well targeted products in each market. This may lead us to erroneously conclude that there is no long tail as described in the existing literature. Simulating our model with the same small number

Figure 2.5: Local Tail: Correcting for Small Samples



of sales in each local market yields very similar results.[10] 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 with 23.6% to 31.6% of sales coming from products outside the top 3,000 products.

Table 2.12: Average Revenue Share of Products Outside of the Top 3,000

	Small Sample		Large Sample	
	Data	Model (Logit)	Logit	Nested
National	50.6%	49.4%	50.6%	50.6%
Local	3.4%	2.6%	31.6	23.6%

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

2.7 Robustness

In this section, we examine the robustness of our findings to the size of the counterfactual choice set. While we find that size of the overstatement is sensitive to the size of the counterfactual assortment size, our findings from the previous section are on the lower end, suggesting that our conclusions are on the conservative side. Tables 2.13 and 2.14 present the change in consumer welfare and the size of the overstatement resulting from various thresholds of the counterfactual choice set for the simple and nested logit specifications, respectively. For comparison, we also include our baseline results from the previous section.

Table 2.13: Robustness: Overstatement of Consumer Welfare Increase (Simple Logit)

	Percent Increase			Absolute ($ Millions)			
Assortment Size	Loc.	Nat.	%Δ	Loc.	Nat.	Δ	%Δ
Baseline	10.4	45.5	337.5	57.69	194.68	136.99	237.46
Threshold							
3,000	42.2	169.9	302.6	182.01	386.37	204.36	112.28
6,000	24.5	98.4	301.6	108.42	304.41	195.99	180.77
12,000	6.1	45.3	642.6	35.28	191.45	156.17	442.66
24,000	0.5	14.6	2920	3.39	78.28	74.89	2209.14

Unsurprisingly, as the size of the counterfactual choice set increases the gain to consumers from access to the remaining products decreases. This decrease occurs substantially faster under locally customized assortments than under nationally standardized assortments. As a result, the percentage overstatement is increasing in the assortment size, despite the absolute size of the overstatement decreasing. This pattern is illustrated in Figure 2.7. Figure 2.7 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). In the simple logit setup, the absolute consumer welfare overstatement peaks at $204.5 with about about 6% of products (or 3,000 shoe varieties) and in the nested logit setup, the absolute overstatement peaks at $207.9 million with

Table 2.14: Robustness: Overstatement of Consumer Welfare Increase (Nested Logit)

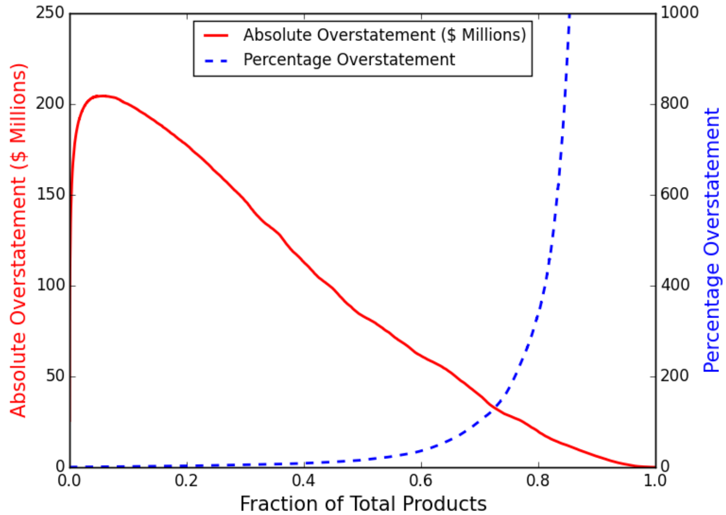
	Percent Increase			Absolute ($ Millions)			
Assortment Size	Loc.	Nat.	%Δ	Loc.	Nat.	Δ	%Δ
Baseline	6.8	19.5	186.8	57.20	146.63	89.43	156.35
Threshold							
3,000	24.6	71.1	189.0	354.19	372.74	18.55	5.24
6,000	13.7	45.6	232.8	216.61	281.11	64.50	29.78
12,000	4.7	19.8	321.3	81.00	148.13	67.13	82.88
24,000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

about about 2% of products (or 1,000 shoe varieties). For both setups, the percentage overstatement approaches infinity as the fraction of available products approaches one.

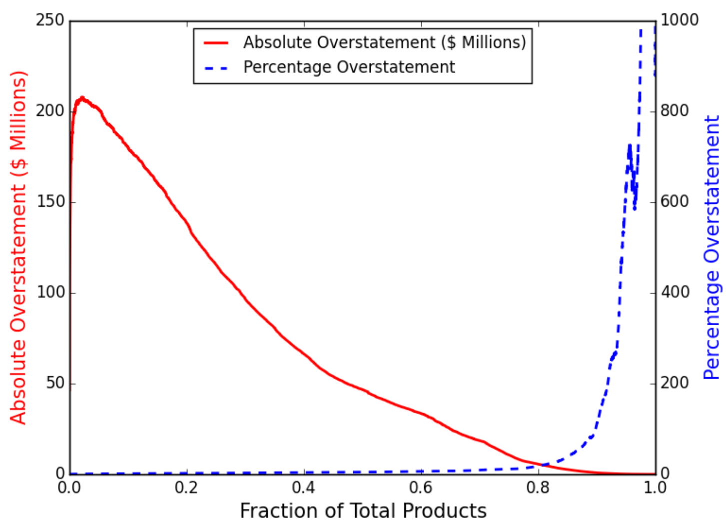
Tables 2.15 and 2.16 present the retail revenue at various thresholds of the counterfactual choice set for simple and nested logit, respectively. With retail revenue, we find that as assortment size increases, the gain from customizing assortments to local demand is decreasing. However, a typical large brick-and-mortar shoe retailer stocks, at most, a few thousand varieties. This puts them at the small end of our robustness analysis, suggesting there may be significant incentives for large national brick-and-mortar shoe retailers to customize their assortments to local demand.

Figure 2.7 graphs the increase in retail revenue due to local customization of assortments, measured in millions of dollars (solid) and as a percentage (dash). In the simple logit setup, the absolute gain in revenue from localization peaks at $273.4 million at about 5% of products (or 2,500 shoe varieties), while in the nested logit setup, the absolute gain in revenue from localization peaks at $165.9 million at about 3% of products (or 1,500 shoe varieties). The percentage gain is monotonically decreasing with assortment size. The graph shows that when assortment sizes are extremely limited, brick-and-mortar retailers can significantly boost revenue by maintaining locally customized product assortments.

Figure 2.6: Overestimation of Consumer Welfare



(a) Simple Logit



(b) Nested Logit

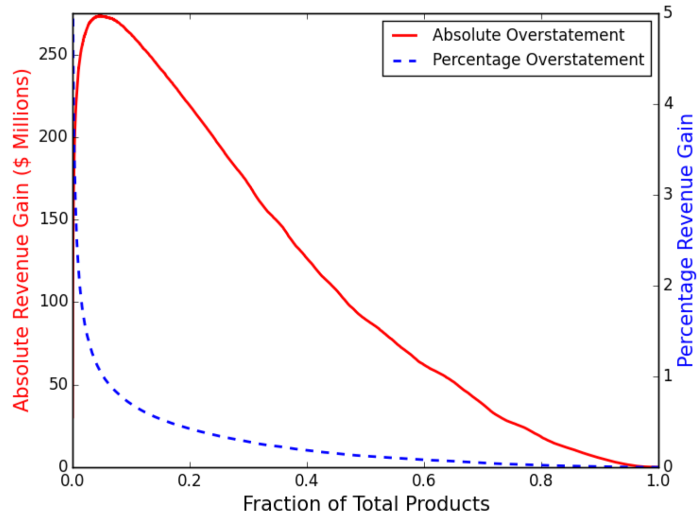
Table 2.15: Robustness: Retail Revenue (Simple Logit)

Assortment Size	Absolute ($ Millions) Loc.	Nat.	Δ	$\%\Delta$
Baseline	742.50	576.53	165.97	28.79
Threshold				
3,000	565.36	292.97	272.39	92.98
6,000	665.44	411.52	253.92	61.70
12,000	771.07	586.03	185.04	31.58
24,000	820.78	742.45	78.33	10.55

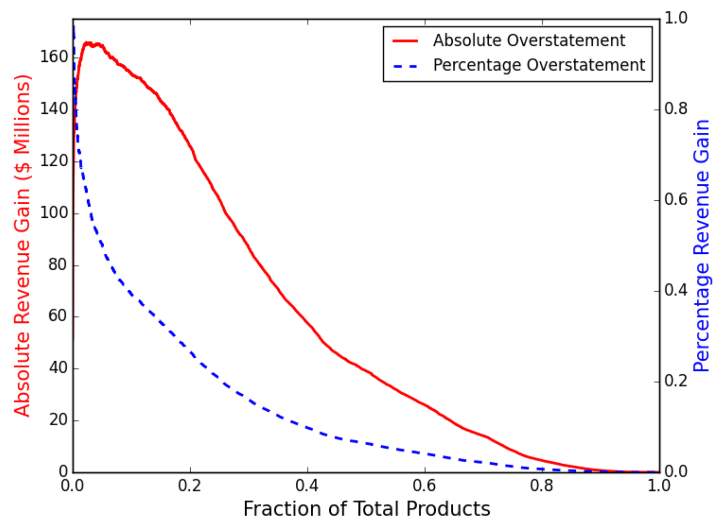
Table 2.16: Robustness: Retail Revenue (Nested Logit)

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

Figure 2.7: Increase in Retail Revenue from Local Assortments



(a) Simple Logit



(b) Nested Logit

2.8 Conclusion

In this chapter, we quantified the effect of increased access to variety due to online retail on consumer welfare and firm profitability. The value of online variety depends on the set of products that would be available through traditional brick-and-mortar retailers in the absence of the internet. Since traditional brick-and-mortar retailers tend to cater their product assortments to local demand, we highlight the importance of accounting for across-market differences in demand and assortments. We built a new micro-level data set containing the sales of footwear by a large online retailer to estimate a rich model of demand allowing for consumer demand heterogeneity across markets.

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

We use new methodology developed in Chapter 1 to confront our small samples problem. Rather than use disaggregated local market shares directly, we use our information on location-specific sales as a type of micro moment to augment our estimation with aggregated sales data. Our estimation strategy exploits the structure of the model to separate the problem into two parts. At the aggregate level our estimation mimics the standard approach to pin down the demand parameters common across locations. Separately, our micro moments are used to estimate the distribution of consumer heterogeneity across markets.

We find products face substantial heterogeneity in demand across markets, with more niche products facing greater heterogeneity. We also show that accounting for this heterogeneity is important for rationalizing the distribution of local sales. Using our estimated model, we run a series of counterfactuals. In this analysis we find that abstracting from across-market demand heterogeneity overestimates the consumer welfare gain due to online markets by up to 338%. On the supply side, our estimates suggest that brick-and-mortar retail chains generate up to 28.8% additional revenue by localizing their assortments. Finally, we revisit the long tail phenomenon in online retail. Our results suggest that inferring consumer welfare gains from the observed aggregate long tail will tend to overstate actual welfare gains because the aggregation of sales over markets with differing demand is a key driver of the long tail.

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

Chapter 3

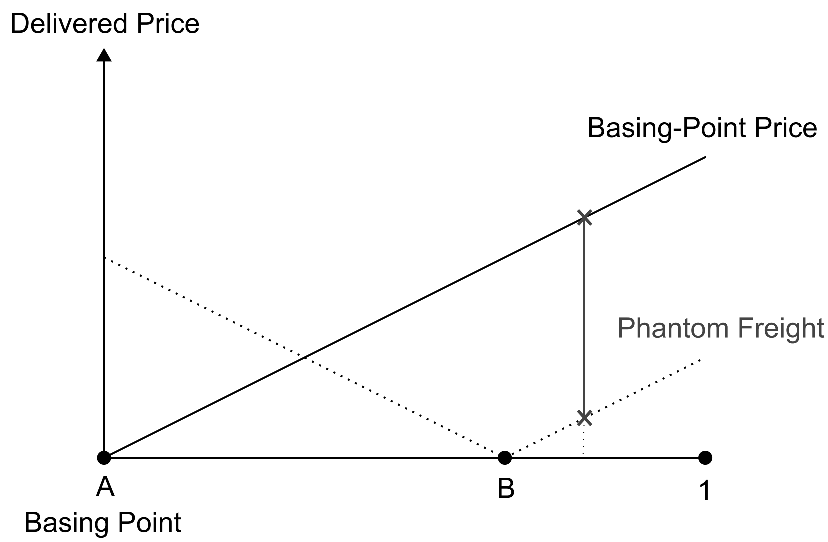
Bertrand Competition and Non-Collusive Basing-Point Pricing

3.1 Introduction

In an environment with spatial differentiation, basing-point pricing occurs when the delivered prices faced by a consumer is determined by the consumer's distance from a common location, known as the basing point. That is, the price faced by a consumer at any given location is equal to the price charged by firms at the basing point plus the cost of transportation from the basing point to the consumer's location regardless of the actual location of the selling firm. The difference in transportation costs, pocketed by the selling firm, is often referred to as phantom freight. Figure 3.1 illustrates this phenomenon for a simple unit line location model.

Much of the early work on the topic, such as Stigler (1949) and McGee (1954), argued that basing-point pricing was a mechanism for firms at different locations to collude. A couple of the more noteworthy real world examples, "Pittsburgh Plus" and the "Eau Claire Rule," are commonly understood to be the result of collusion, the latter being government imposed. In the early 1900's, under the Pittsburgh Plus pricing system, consumers of steel would be charged the price of steel at Pittsburgh plus the cost of

Figure 3.1: Basing-Point Pricing Illustration



Consider an economy of unit length with consumers having inelastic demand uniformly distributed over the line. Suppose production occurs in two locations, sites $A = 0$ and $B \in (0, 1)$. For convenience, assume firms at both sites face constant marginal costs equal to zero and transportation costs, τ, are a linear function of distance. Further suppose that competition at site A results in the competitive mill price equal to zero. Then the basing-point price is the transportation cost from site A to the consumers location. The transportation costs of site B are represented by the dotted lines emanating in either direction from the point B. If firms at site B charge consumers the basing-point price, they obtain phantom freight equal to the difference between site A and site B transportation costs to deliver to that consumer.

transporting that steel from Pittsburgh to the consumer's location regardless of the actual seller's location. The Eau Claire Rule was a federal pricing scheme established in the 1930's to support local dairy farmers. It set minimum prices processors must pay to farmers for beverage-grade milk. The price floor for a given market was set based upon the distance of that market from Eau Claire, Wisconsin (Killman 1997).

From a purely price standpoint, basing-point pricing makes the choice of producer irrelevant to the consumer, resulting in abnormalities, such as freight absorption and cross-hauling. Since transportation is costly, these abnormalities are commonly viewed as inefficient consequences of collusion. Freight absorption occurs when producers in relatively high price areas ship goods toward areas with relatively low prices. In order to make this sale, the producer will absorb some of the transportation cost to make its delivered price competitive with the low price goods. Cross-hauling occurs when goods produced at location A cross goods produced at location B. An extreme example is when producers at location A are shipping to consumers at location B, while simultaneously producers at location B are shipping to consumers at location A.

On the opposing side of the argument, Haddock (1982) formalized a theory of basing-point pricing that was not the result of collusion across production sites. In his model, firms face constant marginal costs, but allows for potential scale economies in transportation. There are multiple producers at the basing point, but he is agnostic about the form of competition, while there is only one monopolistic non-basing point firm. Facing inelastic demand from each location, the non-basing point firm matches the delivered price of the basing point in order to extract spatial rent. He further argued that basing-point pricing is inconsistent with an effective cartel because there almost always exists a more profitable alternative strategy. In particular, a cartel could increase its profits by eliminating the inefficient anomalies discussed in the previous paragraph. Additionally, he put forth several potential competitive motivations for cross-hauling.[1]

The results in Haddock (1982) were later refined by Gilligan (1993). Gilligan notes that basing-point pricing is only consistent with marginal cost pricing at the base site. Otherwise, the delivered price gradient would also reflect spatial price discrimination rather than transportation costs alone. Additionally, he allows the non-basing point

[1] These motivations include preventing supply interruption, information costs, and decreasing marginal transportation cost as a function of distance.

production site to contain more than a single monopoly firm and derives conditions under which a basing-point pricing equilibrium exists. He concludes that if non-basing point firms are von Stackleberg leaders with respect to basing-point firms, basing-point pricing can arise without perfect collusion if demand in their markets are large relative to unit transportation costs and the number of firms.

In this chapter, I extend the non-collusive theories of basing-point pricing to allow for Bertrand competition among firms located outside of the basing point.[2] When non-basing point firms face increasing marginal costs that rise at a sufficiently high rate, these firms stop competing amongst each other and instead compete against the basing-point. Under Bertrand competition, non-basing point firms match the delivered price of the basing point resulting in a basing-point pricing pattern.

This model has several attractive features. First, this model allows cross-hauling to arise without imposing economies in long-haul transportation or product heterogeneity. Depending on the distances involved in a particular application, the imposition of long-haul scale economies may not be a plausible assumption. Additionally, in many real world examples of basing-point pricing the products involved were highly, if not perfectly, substitutable, potentially making the product heterogeneity explanation unattractive. Second, the basing-point pricing result can easily be extended to account for multiple non-basing points. In particular, adding a second non-basing point production site would not necessarily cause the basing-point price result to breakdown in my setup. Third, if production at a non-basing point site grows sufficiently large, that site would itself become an additional basing point. Further, I extend the model from a unit line to a circular setting with the basing point in the center. An interesting implication of the additional directional dimension is that basing-point pricing may be observed in some radial markets, while failing in others.

This chapter will proceed as follows. The next section presents the unit line model. Section 3 derives the basing-point pricing result. A discussion of the implications of the model and results can be found in section 4. Section 5 extends the model and the basing-point pricing result to a circular environment and Section 6 concludes the paper.

[2] Using his model with constant marginal costs of production, Gilligan (1993) notes that "Bertrand behavior by non-base site firms is never consistent with basing-point pricing."

3.2 Model

Consider an economy of unit length with location indexed by $r \in [0,1]$. The economy is populated by identical consumers distributed uniformly over the line, each having inelastic demand x. Production of an identical consumption good occurs at two sites, A and B. Site A is situated at location 0, A=0, and site B is located at some point between site A and the end of the line, $B \in (0,1)$. At site A, $N_A > 1$ firms face equal and constant marginal costs, while at site B, N_B firms face increasing marginal costs. Per unit transportation costs vary linearly with distance from the purchasing site,

$$\tau(r,s) = \tau|r-s|, \forall r \in [0,1], s \in \{A,B\}.$$

Under Bertrand competition, profits for a firm at site A, $a \in \{1,...,N_A\}$, are

$$\pi_a = \int_0^1 p_a(r)q_a(r)dr - C_A(Q_a),$$

where $p_a(r)$ is the mill price firm a charges to location r, $q_a(r)$ is the quantity firm a sells to location r, and $C_A(\cdot)$ is the cost of producing $Q_a = \int_0^1 q_a(r)dr$ units of the consumption good at site A. Similarly, profits for a firm at site B, $b \in \{1,...,N_B\}$, are

$$\pi_b = \int_0^1 p_b(r)q_b(r)dr - C_B(Q_b).$$

Define the delivered price offered to location r to be the sum of the mill price charged to location r and the cost of transportation to location r,

$$\begin{aligned} \hat{p}_a(r) &= p_a(r) + \tau r \\ \hat{p}_b(r) &= p_b(r) + \tau|r-B|, \end{aligned}$$

then the quantity demanded at location r from a site A firm is

$$q_a(r) = \begin{cases} 0 & \text{if } \exists a' \text{ s.t } \hat{p}_a(r) > \hat{p}_{a'}(r) \text{ or } \exists b \text{ s.t } \hat{p}_a(r) > \hat{p}_b(r) \\ \frac{x}{N_A} & \text{if } \hat{p}_a(r) = \hat{p}_{a'}(r), \forall a' \text{ and } \hat{p}_a(r) < \hat{p}_b(r), \forall b \\ x & \text{if } \hat{p}_a(r) < \hat{p}_{a'}(r), \forall a' \text{ and } \hat{p}_a(r) < \hat{p}_b(r), \forall b. \end{cases}$$

and from a site B firm

$$q_b(r) = \begin{cases} 0 & \text{if } \exists b' \text{ s.t } \hat{p}_b(r) > \hat{p}_{b'}(r) \text{ or } \exists a \text{ s.t } \hat{p}_b(r) > \hat{p}_a(r) \\ \frac{x}{N_B} & \text{if } \hat{p}_b(r) = \hat{p}_{b'}(r), \forall b' \text{ and } \hat{p}_b(r) \le \hat{p}_a(r), \forall a \\ x & \text{if } \hat{p}_b(r) < \hat{p}_{b'}(r), \forall b' \text{ and } \hat{p}_b(r) \le \hat{p}_a(r), \forall a. \end{cases}$$

Definition. An equilibrium is a set of delivered prices $\{\hat{p}_a^*, \hat{p}_b^*\} = \{\hat{p}_a^*(r), \hat{p}_b^*(r)\}_{r\in[0,1]}$ and quantities $\{q_a^*, q_b^*\} = \{q_a^*(r), q_b^*(r)\}_{r\in[0,1]}$, $a \in \{1,...,N_A\}, b \in \{1,...,N_B\}$, s.t.

1. A firm, $s \in \{1,...,N_S\}$, at site $S \in \{A,B\}$, chooses prices, $\hat{p}_s$, to maximize profits, given prices of other firms, $\hat{p}_{-s}^*$, i.e.

$$\pi_s(\hat{p}_s^*; \hat{p}_{-s}^*) \geq \pi_s(\hat{p}_s; \hat{p}_{-s}^*)$$

2. No arbitrage: The prices of a firm, $s \in \{1,...,N_S\}$, at site $S \in \{A,B\}$ satisfy

$$\hat{p}_s^*(r) \leq \hat{p}_s^*(r') + \tau|r - r'|, \quad \forall r, r' \in [0,1].$$

3. Markets clear:

$$\sum_a^{N_A} q_a^*(r) + \sum_b^{N_B} q_b^*(r) = x, \quad \forall r \in [0,1].$$

3.3 Basing-Point Pricing

Under Bertrand competition, competition at site A results in a mill price equal to marginal cost, $p_a(r) = p_A(r) = C_A'$, $\forall a$, and, for convenience, assume the that $C_A' = 0$. A consumer at location $r \in [0,1]$ can then purchase the consumption good from any site A firm, $a \in \{1,...,N_A\}$ at the price

$$\hat{p}_A(r) = \tau r.$$

The following lemma shows that under basing-point pricing firms at site B will satisfy all demand to the right of site B before selling to the left, toward site A.

Lemma 1. Under basing-point pricing, firms at site B strictly prefer to sell to any point to their right, $[B,1]$, over selling to any point to their left, $[0,B)$.

Proof. The difference in transportation costs between site A and site B of delivering to location r is

$$\tau(r,A) - \tau(r,B) = \begin{cases} \tau r - \tau(B - r) = 2\tau r - \tau B & \text{if } r < B \\ \tau r - \tau(r - B) = \tau B & \text{if } r \geq B. \end{cases}$$

For any $r < B$, $2r - B < B$, that is, the difference in transportation costs between site A and site B of delivering to location r is greater to the right of site B. Therefore, under basing-point pricing, the marginal revenue of selling to locations to the right of B is greater than marginal revenue of selling to the left of B. ■

Define $q_B(r) = \sum_{b=1}^{N_B} q_b(r)$. The following proposition establishes conditions for the basing-point pricing property under Bertrand competition. The key to the result is that marginal costs for site B firms rise sufficiently quickly so that it is unprofitable to sell to any consumers to the left of site B. In this situation, firms at site B are able to sell as much of the good as they please at the site A basing-point price and, as a result, do not compete against other site B firms.

Proposition 1. Suppose $C_B(\cdot)$ is a continuous function s.t. $C_B(0) = 0, C'_B(\cdot) \geq 0$, and $C''_B(\cdot) > 0$. If

$$C'_B\left(\int_B^1 q_B(r)dr\right) \geq \hat{p}^*_A(1) - \tau(1-B) = \tau B$$

then $d\hat{p}^*_A(r) = d\hat{p}^*_B(r) = \tau$, $\forall r$.

Proof. Due to symmetry, all site B firms behave alike, so $\hat{p}^*_b = \hat{p}^*_B$, $\forall b$. Since competition at site A results in

$$\hat{p}^*_A(r) = \tau r, \quad \forall r \in [0,1]$$

and

$$\frac{d\hat{p}^*_A(r)}{dr} = \tau,$$

it suffices to show that $\hat{p}^*_B(r) = \hat{p}^*_A(r)$ for any $r \in [0,1]$ s.t. $q_b(r) > 0$.

By Lemma 1, we know that firms at site B will first satisfy demand to their right, $[B,1]$. The conditions on site B's cost function suggest that site B firms will fall short of satisfying or just satisfy demand in $[B,1]$. Thus there will be no site B sales to the left of site B, $[0,B)$. With this knowledge, I begin by finding the site B's optimal level of production at the basing-point price, then show that no site B firms want to deviate from basing-point pricing.

Note that if $C'_B(0) > \hat{p}^*_A(B)$, the result is trivial because site B firms will make no sales. Thus, suppose $C'_B(0) \leq \hat{p}^*_A(B)$. Then by the intermediate value theorem

$\exists \bar{B} \in [B, 1]$ be s.t.

$$C_B'\left(\int_B^{\bar{B}} q_B(r)dr\right) = \hat{p}_A^*(\bar{B}) - \tau(\bar{B} - B) = \tau B$$

I show that it is unprofitable for a site B firm, $b \in \{1, ..., N_B\}$, to deviate from $\hat{p}_b = \hat{p}_B^* = \hat{p}_A^*$ and

$$q_b^*(r) = q_B^*(r) = \begin{cases} \frac{x}{N_B} & \text{if } r \in [B, \bar{B}] \\ 0 & \text{else} \end{cases}.$$

The profit of firm b at the basing-point price is

$$\begin{aligned}
\pi_b(\hat{p}_A^*) &= \frac{1}{N_B}\int_B^{\bar{B}} [\hat{p}_A^*(r)x - \tau(r - B)x]dr - C_B\left(\frac{1}{N_B}\int_B^{\bar{B}} xdr\right) \\
&= \frac{1}{N_B}\int_B^{\bar{B}} \tau Bxdr - C_B\left(\frac{1}{N_B}\int_B^{\bar{B}} xdr\right) \\
&= \frac{\tau Bx(\bar{B} - B)}{N_B} - C_B\left(\frac{x(\bar{B} - B)}{N_B}\right) \\
&= \tau BQ_b(\hat{p}_A^*) - C_B(Q_b(\hat{p}_A^*)).
\end{aligned}$$

At the basing-point price the marginal profit of a change in price at some location $r \in [0, 1]$ is equal to

$$\left.\frac{\partial \pi_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} = \begin{cases} \left[\tau B - C_B'(Q_b(\hat{p}_b))\right] \left.\frac{\partial Q_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} & \text{if } r \in [B, 1] \\ \left[2\tau r - \tau B - C_B'(Q_b(\hat{p}_b))\right] \left.\frac{\partial Q_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} & \text{if } r \in [0, B) \end{cases}$$

Note that

$$\begin{aligned}
&\left.\frac{\partial Q_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} > 0 \text{ if } \partial p_b(r) < 0 \\
&\left.\frac{\partial Q_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} < 0 \text{ if } \partial p_b(r) > 0 \text{ and } r \in [B, \bar{B}] \\
&\left.\frac{\partial Q_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b = \hat{p}_A^*} = 0 \text{ if } \partial p_b(r) > 0 \text{ and } r \in [0, B) \cup (\bar{B}, 1]
\end{aligned}$$

Additionally, since $C_B'(Q_b(\hat{p}_A^*)) = \tau B$ and $C''(\cdot) > 0$,

$$\begin{aligned}
C_B'(Q_b(\hat{p}_b)) &\geq \tau B \text{ if } \partial p_b(r) < 0 \\
C_B'(Q_b(\hat{p}_b)) &\leq \tau B \text{ if } \partial p_b(r) > 0.
\end{aligned}$$

Therefore,

$$\left.\frac{\partial \pi_b(\hat{p}_b)}{\partial \hat{p}_b(r)}\right|_{\hat{p}_b=\hat{p}^*_A} \leq 0,$$

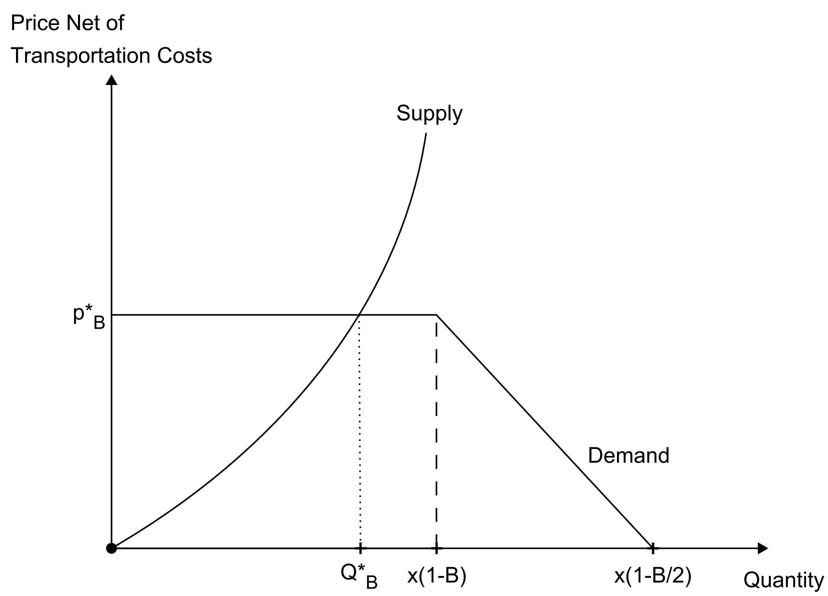
i.e.

$$\pi^*_b(\hat{p}^*_A) \geq \pi^*_b(\hat{p}_b),$$

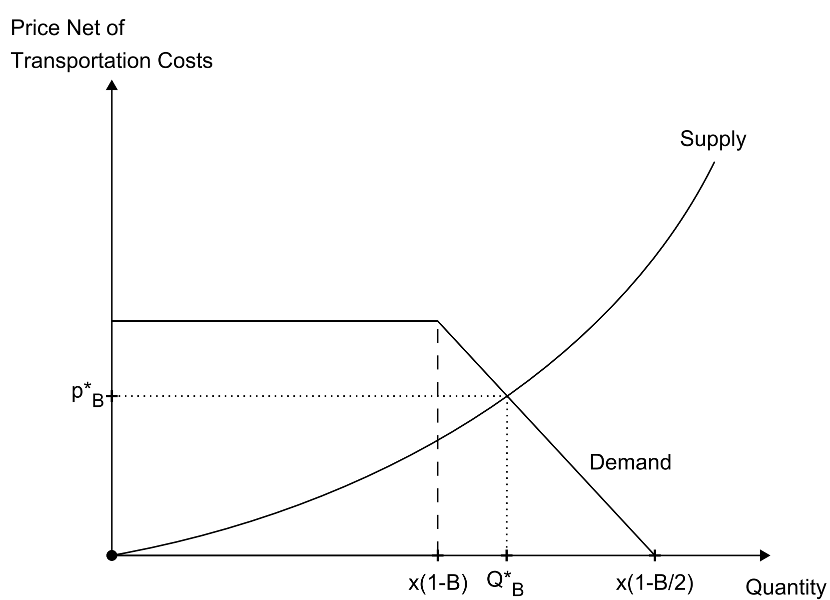
for any alternative vector of delivered prices $\hat{p}_b$. ■

Figure 3.2(a) illustrates this result. This figure plots the supply and demand curves faced by site B as a function marginal revenue, which is equal to the delivered price net of transaction costs. The demand curve faced by site B begins flat because the difference in transportation costs between sites A and B is constant at all points to the right of site B. Net of transportation costs, consumers to the right of site B are willing to pay up to $p^*_B(r) = \tau B$. With $(1 - B)$ consumers, each having inelastic demand of x, to the right of site B, site B firms can sell up to $Q_B = x(1-B)$ units of the consumption good at this price. In order to increase above $Q_B = x(1-B)$, site B firms would have to turn to consumers to the left of site B, toward site A. In this direction, transportation costs from site A are decreasing, while transportation costs from site B are increasing. As a result, site B firms will have to charge lower prices to sell to these consumers. After $Q_B = x(1 - B/2)$, the midpoint between A and B, transportation costs from site B are greater than the delivered price of site A. Thus, selling to consumers beyond this point would be unprofitable.

If site B's aggregate marginal cost (supply) curve crosses the demand curve before $Q_B = x(1-B)$, the competitive Bertrand equilibrium outcome is for all site B firms to charge a delivered price equal to the basing-point price, $\hat{p}^*_B(r) = \hat{p}^*_A(r) = \tau r$, or a price net of transaction costs equal to $p^*_B(r) = \tau B$. As a result, site B firms do not compete against each other, but only against site A. Intuitively, this is because site B firms are not concerned about being undercut by other site B firms. (1) If a site B firm were to be undercut it knows that it could find another consumer willing to pay the basing-point price. (2) Every site B firm is able to sell as much as it pleases at the basing-point price, attempting to increase its market share by undercutting another site B firm would only serve to increase its marginal costs. On the other hand, Figure 3.2(b) illustrates an equilibrium in which site B's aggregate marginal cost curve intersects the demand curve after $Q_B = x(1-B)$. In this case, site B firms will compete against each other and the



(a) Basing-Point Pricing Equilibrium



(b) Non-Basing-Point Pricing Equilibrium

Figure 3.2: Site B - Supply and Demand

market price charged by site B firms will be below the basing-point price. This result is formalized in the following proposition.

Proposition 2. Suppose $C_B(0) = 0, C'_B(\cdot) \geq 0$, and $C''_B(\cdot) > 0$. If

$$C'_B\left(\int_B^1 q_B(r)dr\right) < \tau B$$

then $d\hat{p}^*_B(r) \neq \tau$, $\forall r$.

Proof. By the intermediate value theorem $\exists \underline{B} \in \left[\frac{B}{2}, B\right]$ s.t.

$$C'_B\left(\int_{\underline{B}}^1 q_B(r)dr\right) = \hat{p}^*_A(\underline{B}) - \tau(B - \underline{B}) = \tau(2\underline{B} - B).$$

It is straightforward to show that in equilibrium, site B firms charge a mill price equal to

$$p^*(r) = \tau(2\underline{B} - B).$$

As a result,

$$d\hat{p}^*_B(r) = \begin{cases} -\tau & \text{if } r \in [0, B) \\ \tau & \text{if } r \in [B, 1] \end{cases},$$

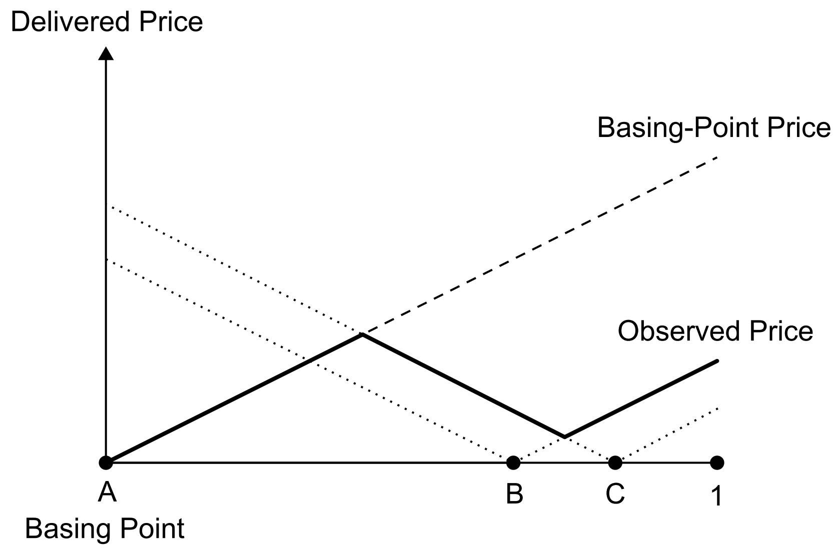
that is basing-point pricing does not occur in equilibrium. ■

3.4 Implications

The results present a clear and testable implication of my theory for competitive basing-point pricing stemming from Lemma 1. It implies that non-basing point firms only sell to the right, away from the basing point. Additionally, the model has several attractive features. First, notice that the condition for basing-point pricing is that non-basing point firms are unable to fully satisfy demand to their right. As a consequence, sales from site A are required to satisfy the residual demand. It follows that, in this model, cross-hauling occurs without needing to impose economies of long-haul transportation or product heterogeneity.

Second, a common criticism of current non-collusive theories of basing-point pricing is that the result breaks down with the addition of another, well placed, non-basing point.[3] Suppose a second non-basing point site, site C, is added to the model above, between site B and the end of the line. Under the current theories of non-collusive basing-point pricing, firms at sites B and C will directly compete with each other. Thus, if site B attempted to charge the basing-point price, site C would profit by undercutting site B and vice versa. As a result, basing-point pricing fails and the pricing pattern for sites B and C is reminiscent of a standard Hotelling model with locational price discrimination. This example is illustrated in Figure 3.3.

Figure 3.3: Multiple Non-Basing Points



On the other hand, it is straightforward to extend the basing-point pricing result with an additional non-basing point in my model. First, one would show that proposition 1 holds for site C, then show that it holds for site B with the residual demand left over by site C. In this way, the basing-point pricing result can be further generalized to an arbitrary number of non-basing-point sites.

Third, it has been observed that industries that develop a basing-point pricing system often begin with a large production center, typically of historic importance, as the basing point. However, when production at a non-basing point location grows to a significant level, that location itself will become an additonal basing point, transforming

[3] See Benson, Greenhut, and Norman (1990).

the pricing scheme into one with multiple basing points. My model allows for this possibility. In the example above with three production sites, suppose production at site C grew sufficiently large to mimic constant marginal costs, i.e. the number of site C firms approached infinity, $N_C \to \infty$. The economy would then have two basing points, sites A and C, with site B charging either the site A or site C delivered price depending on the consumer's location.

3.5 Circular Extension

Consider a circular economy with radius $R = 1$ and denote locations by their polar coordinates (r, ϕ). The economy is populated by identical consumers distributed uniformly over the circle, each having inelastic demand x. Production of the consumption good is produced at two sites, A and B. Site A is situated at the center of the circle, A=0, and site B is a ring located at some radius between site A and the border of the circle, $B = \{(r_B, \phi_B)\}_{\phi_B \in [0,2\pi]}$, where $r_B \in (A, 1)$. Locations at site B are then identified by their angle ϕ_B. At site A, $N_A > 1$ firms face equal and constant marginal costs, while locations at site B, $\{N_{\phi_B}\}_{\phi_B \in [0,2\pi]}$ firms face increasing marginal costs. Per unit transportation costs vary linearly with distance from the purchasing site,

$$\tau((r,\phi), A) = \tau r, \forall r \in [0,1],$$
$$\tau((r,\phi),(r_B,\phi_b)) = \tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}}, \forall r \in [0,1], b \in B.$$

Under Bertrand competition, profits for a firm at site A, $a \in \{1, ..., N_A\}$, are

$$\pi_a = \int_0^1 \int_0^{2\pi} \left[p_a(r,\phi) - \tau r\right] q_a(r,\phi) d\phi dr - C_A(Q_a),$$

where $p_a(r,\phi)$ is the mill price firm a charges to location (r,ϕ), $q_a(r,\phi)$ is the quantity firm a sells to location (r,ϕ), and $C_A(\cdot)$ is the cost of producing $Q_a = \int_0^1 \int_0^{2\pi} q_a(r,\phi) d\phi dr$ units of the consumption good at site A. Similarly, profits for a firm at site B, (b, ϕ_b), $b \in \{1, ..., N_{\phi_b}\}$, are

$$\pi_b = \int_0^1 \int_0^{2\pi} \left[p_b(r,\phi) - \tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}}\right] q_b(r,\phi) d\phi dr - C_B(Q_b).$$

Define the delivered price offered to location r to be the sum of the mill price charged to location r and the cost of transportation to location r,

$$\hat{p}_a(r,\phi) = p_a(r) + \tau r$$
$$\hat{p}_b(r,\phi) = p_b(r,\phi) + \tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}},$$

then the quantity demanded at location r from a site A firm is

$$q_a(r,\phi) = \begin{cases} 0 & \text{if } \exists a' \text{ s.t } \hat{p}_a(r,\phi) > \hat{p}_{a'}(r,\phi) \text{ or } \exists b \text{ s.t } \hat{p}_a(r,\phi) > \hat{p}_b(r,\phi) \\ \frac{x}{N_A} & \text{if } \hat{p}_a(r,\phi) = \hat{p}_{a'}(r,\phi), \forall a' \text{ and } \hat{p}_a(r,\phi) < \hat{p}_b(r), \forall b \\ x & \text{if } \hat{p}_a(r,\phi) < \hat{p}_{a'}(r,\phi), \forall a' \text{ and } \hat{p}_a(r,\phi) < \hat{p}_b(r,\phi), \forall b. \end{cases}$$

and from a site B firm

$$q_b(r,\phi) = \begin{cases} 0 & \text{if } \exists b' \text{ s.t } \hat{p}_b(r,\phi) > \hat{p}_{b'}(r,\phi) \text{ or } \exists a \text{ s.t } \hat{p}_b(r,\phi) > \hat{p}_a(r,\phi) \\ \frac{x}{N_B} & \text{if } \hat{p}_b(r,\phi) = \hat{p}_{b'}(r,\phi), \forall b' \text{ and } \hat{p}_b(r,\phi) \leq \hat{p}_a(r,\phi), \forall a \\ x & \text{if } \hat{p}_b(r,\phi) < \hat{p}_{b'}(r,\phi), \forall b' \text{ and } \hat{p}_b(r,\phi) \leq \hat{p}_a(r,\phi), \forall a. \end{cases}$$

Definition. An equilibrium is a set of delivered prices $\{\hat{p}_a^*, \hat{p}_b^*\} = \{\hat{p}_a^*(r,\phi), \hat{p}_b^*(r,\phi)\}_{r\in[0,1],\phi\in[0.2\pi]}$ and quantities $\{q_a^*, q_b^*\} = \{q_a^*(r,\phi), q_b^*(r,\phi)\}_{r\in[0,1],\phi\in[0,2\pi]}$, $a \in \{1,...,N_A\}, b \in \{1,...,N_{\phi_B}\}_{\phi_B\in[0,2\pi]}$, s.t.

1. A firm, $s \in \{1,...,N_S\}$, at site $S \in \{A, \phi_B\}$, chooses prices, $\hat{p}_s$, to maximize profits, given prices of other firms, $\hat{p}_{-s}^*$, i.e.

$$\pi_s(\hat{p}_s^*;\hat{p}_{-s}^*) \geq \pi_s(\hat{p}_s;\hat{p}_{-s}^*)$$

2. No arbitrage: The prices of a firm, $s \in \{1,...,N_S\}$, at site $S \in \{A, \phi_B\}$ satisfy

$$\hat{p}_s^*(r) \leq \hat{p}_s^*(r') + \tau\left(r^2 + r'^2 - 2rr'\cos(\phi,\phi')\right)^{\frac{1}{2}}, \quad \forall r, r' \in [0,1], \phi, \phi' \in [0, 2\pi].$$

3. Markets clear:

$$\sum_a^{N_A} q_a^*(r,\phi) + \int_0^{2\pi} \sum_b^{N_{\phi_B}} q_b^*(r,\phi) d\phi_B = x, \quad \forall r \in [0,1], \phi \in [0,2\pi].$$

The following lemma is analogous to lemma 1. It shows that under basing-point pricing firms at site B will satisfy all demand outward from the center along their ray, ϕ_b, before selling inward toward site A or along any other ray.

Lemma 2. Under basing-point pricing, firms at site B, more specifically located at (r_B, ϕ_b), strictly prefer to sell to any point outward from the center along their ray, $\{(r, \phi_b)\}_{r\in[r_B,1]}$, over selling to any point inward toward the center, $\{(r, \phi_b)\}_{r\in[0,r_B)}$, or along any other ray, $\{(r, \phi)\}_{r\in[0,1],\phi\in[0,2\pi]\setminus\phi_b}$.

Proof. The difference in transportation costs between site A and site B of delivering to location (r, ϕ) is

$$\tau((r,\phi),A) - \tau((r,\phi),(r_B,\phi_b)) = \begin{cases} \tau r - \tau(r_B - r) = 2\tau r - \tau r_B & \text{if } r < B \text{ and } \phi = \phi_b \\ \tau r - \tau(r - r_B) = \tau r_B & \text{if } r \geq B \text{ and } \phi = \phi_b \\ \tau r - \tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}} & \text{if } \phi \neq \phi_b \end{cases}$$

For any $r < r_B$, $2r - r_B < r_B$, that is, the difference in transportation costs between site A and site (r_B, ϕ_b) of delivering to location (r, ϕ_b) is greater outward away from site A. Also, by the triangle inequality, $\tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}} > \tau(r - r_B)$, hence, $\tau r_B > \tau r - \tau\left(r^2 + r_B^2 - 2rr_B\cos(\phi,\phi_b)\right)^{\frac{1}{2}}$. That is, the difference in transportation costs between site A and site (r_B, ϕ_b) of delivering to any location (r, ϕ), $\phi \neq \phi_b$ is smaller than for any location $(r, \phi_b), r \in [r_B, 1]$. Therefore, under basing-point pricing, the marginal revenue of selling to locations outward from the center along its own ray is greater than marginal revenue of selling inward toward the center or any point along another ray. ■

I now state the analogous basing-point pricing result for the circular model. The proof is the same as the proving the linear case individually for each ray $\phi_b \in [0, 2\pi]$.

Proposition 3. Suppose $C_B(\cdot)$ is a continuous function s.t. $C_B(0) = 0, C_B'(\cdot) \geq 0$, and $C_0''(\cdot) > 0$. If

$$C_B'\left(\int_{r_B}^1 q_B(r,\phi_b)dr\right) \geq \tau r_B, \quad \forall \phi_b \in [0, 2\pi]$$

then $\frac{\partial \hat{p}_A^*(r,\phi)}{\partial r} = \frac{\partial \hat{p}_b^*(r,\phi)}{\partial r} = \tau$, $\forall r \in [0,1], \phi \in [0, 2\pi]$.

An interesting implication of the circular model is that if basing-point pricing breaks down in some markets, it need not break down industry-wide. One way to think of this more intuitively is to reconsider the unit line model with the line extended to the left of the basing point, site A, with its own production site, call it site C. Due to transportation costs and with site A in the middle, firms at site B and C cannot compete directly with each other. Since site B and C are separated, the behavior of firms at site C does not affect site B firms and vice versa. Therefore, a breakdown of basing-point pricing to the left of the basing point does not preclude basing-point pricing to the right of the basing-point. The following proposition formalizes this result.

Proposition 4. Suppose that $\exists! \phi_b \in [0, 2\pi]$ s.t.

$$C'_B \left(\int_{r_B}^{1} q_B(r, \phi_b) dr \right) < \tau r_B.$$

Then there $\exists \theta < \pi$ s.t. $\forall \phi \in [0, 2\pi] \backslash [\phi_b + \theta, \phi_b - \theta]$, $\frac{\partial \hat{p}^*_A(r,\phi)}{\partial r} = \frac{\partial \hat{p}^*_b(r,\phi)}{\partial r} = \tau$, $\forall r \in [0, 1]$.

Proof. As in the unit line model, the site B firms at ϕ_b will choose to sell inwards toward site A, but with the additional dimension, they could also sell to adjacent rays. In fact, firms at (r_B, ϕ_b) would be indifferent between selling to consumers at $(r_{\underline{B}}, \phi_b)$, $r_{\underline{B}} \in \left[\frac{r_B}{2}, r_B\right)$ or to consumers at any location (r, ϕ) s.t. the difference in the distance from site A and (r_B, ϕ_b) is $2r_B - r_{\underline{B}}$. That is, competition among firms at (r_B, ϕ_b) will spill over into adjacent rays, potentially putting downward pressure on the prices at these locations.

Denote the distance between two polar coordinates as

$$\triangle((r_1, \phi_1), (r_2, \phi_2)) = \left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\phi_1, \phi_2)\right)^{\frac{1}{2}},$$

and define $\theta < \pi$ s.t.

$$\triangle((1, \phi_b + \theta), A) - \triangle((1, \phi_b + \theta), (r_B, \phi_b)) = 0.$$

That is, site (r_B, ϕ_b) is the same distance from $(1, \phi_b + \theta)$ as site A. Thus, at the basing-point price, the marginal revenue of site (r_B, ϕ_b) firms of selling to $(1, \phi_b + \theta)$ would be zero.

Therefore, by symmetry, $\phi \in [\phi_b - \theta, \phi_b + \theta]$ are the bounds within which competition from site (r_B, ϕ_b) can drive down prices. That is, the breakdown in basing-point pricing is confined within these radial markets and basing-point pricing will prevail in all radial markets $\phi \in [0, 2\pi] \backslash [\phi_b + \theta, \phi_b - \theta]$. ■

3.6 Conclusion

In this paper I have presented a theory of basing-point pricing emerging from Bertrand competition at all production sites. When firms located outside of the basing point site face sufficiently increasing marginal costs (or sufficiently high quantity demand), these firms will compete against the basing point instead of amongst each other. Then, under Bertrand competition, non-basing point firms match the delivered price of firms at the basing point resulting in a basing-point pricing system.

This model has several attractive features not fully embodied by other models of non-collusive basing-point pricing. First, this model allows cross-hauling to arise without imposing economies in long-haul transportation or product heterogeneity. Second, the basing-point pricing result does not break down with the addition of non-basing point production sites and is easily extended to account for multiple non-basing points. Third, if production at a non-basing point site grows sufficiently large, that site would itself become an additional basing point. Finally, I have extended the model from the typical unit line setting to a circular environment, which allows for some firms to follow basing-point pricing, while others do not.

References

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

Benson, B. L., M. L. Greenhut, and G. Norman (1990): "On the Basing-Point System," *American Economic Review*, 80(3), 584–88.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gilligan, T. W. (1993): "Imperfect Competition and Basing-Point Pricing," *Economic Inquiry*, 31(3), 394–409.

Haddock, D. D. (1982): "Basing-Point Pricing: Competitive vs. Collusive Theories," *American Economic Review*, 72(3), 289–306.

Killman, S. (1997): "Byzantine Method of Pricing Milk Won't Be Simplified Anytime Soon," *Wall Street Journal*.

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

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

McGee, J. S. (1954): "Cross Hauling — A Symptom of Incomplete Collusion under Basing-Point Systems," *Southern Economic Journal*, 20(4), 369–379.

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

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

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

Stigler, G. J. (1949): "A Theory of Delivered Price Systems," *American Economic Review*, 39(6), 1144–1159.

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

Verboven, F. (1996): "International Price Discrimination in the European Car Market," *The RAND Journal of Economics*, 27(2), pp. 240–268.

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

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

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

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

Appendix A

Appendix to Chapter 1

A.1 Multiple Levels of Nests

Our technique can be extended to handle multiple levels of nests for further flexibility. Lets begin by deriving the model with two levels. That is, products can be categorized (or grouped) into mutually exclusive and exhaustive sets $c_1 \in C_1 \cup \{0\}$ and products within each set c_1 can be further categorized into mutually exclusive and exhaustive subsets $c_2 \in C_2(c_1)$. Verboven (1996) shows that (local) market shares can be written as,

$$\begin{aligned}\pi_{\ell j} &= Pr_\ell\{\ c_1\ \} \cdot Pr_\ell\{\ c_2 \mid c_1\ \} \cdot Pr_\ell\{\ j \mid c_2, c_1\ \} = \pi_{\ell c_1}\pi_{\ell c_2|c_1}\pi_{\ell j|c_2 c_1} \\ &= \frac{\exp\{I_{\ell c_1}\}}{1+\exp\{I_{\ell c_1}\}} \cdot \frac{\exp\{I_{\ell c_2 c_1}/(1-\lambda_1)\}}{\exp\{I_{\ell c_1}/(1-\lambda_1)\}} \cdot \frac{\exp\{\delta_{\ell j}/(1-\lambda_2)\}}{\exp\{I_{\ell c_2 c_1}/(1-\lambda_2)\}},\end{aligned}$$

where

$$\begin{aligned}I_{\ell c_1} &= (1-\lambda_1)\log\left(\sum_{c_2 \in C_2(c_1)} \exp\left\{\frac{I_{\ell c_2 c_1}}{1-\lambda_1}\right\}\right) \\ I_{\ell c_2 c_1} &= (1-\lambda_2)\log\left(\sum_{j \in c_2} \exp\left\{\frac{\delta_j + \eta_{\ell j}}{1-\lambda_2}\right\}\right).\end{aligned}$$

Define

$$D_{\ell c_2} = \sum_{j \in c_2} \exp\left\{\frac{\delta_j + \eta_{\ell j}}{1-\lambda_2}\right\}$$
$$D_{\ell c_1} = \sum_{c_2 \in C_2(c_1)} D_{\ell c_2}^{\frac{1-\lambda_2}{1-\lambda_1}},$$

then we can rewrite

$$I_{\ell c_1} = (1-\lambda_1)\log(D_{\ell c_1})$$
$$I_{\ell c_2 c_1} = (1-\lambda_2)\log(D_{\ell c_2})$$

and we have

$$\pi_{\ell c_1} = \frac{D_{\ell c_1}^{1-\lambda_1}}{1+\sum_{c_1 \in C_1} D_{\ell c_1}^{1-\lambda_1}}$$
$$\pi_{\ell c_2|c_1} = \frac{D_{\ell c_2}^{(1-\lambda_2)/(1-\lambda_1)}}{D_{\ell c_1}}$$
$$\pi_{\ell j|c_2,c_1} = \frac{\exp\{(\delta_j + \eta_{\ell j})/(1-\lambda_2)\}}{D_{\ell c_2}}.$$

Proposition 3 (Nested Logit - Two Levels)**.** *For any set of* $\{\eta_\ell\}_{\ell=1}^{L}$ *the market share inversion takes the following analytic form,* $\forall j \in J$*,*

$$\delta_j = (1-\lambda_2)\left(\log(\pi_j) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda_2}\right\}\right)\right). \tag{A.1}$$

Proof. By Bayes rule

$$\pi_j(\eta_\ell;\delta,\lambda) = Pr_\ell\{\ c_1\ \} \cdot Pr_\ell\{\ c_2 \mid c_1\ \} \cdot Pr_\ell\{\ j \mid c_2, c_1\ \}$$
$$= \pi_{\ell c_1} \cdot \pi_{\ell c_2|c_1} \cdot \frac{\exp\{(\delta_j + \eta_{\ell j})/(1-\lambda_2)\}}{D_{\ell c_2}},$$

Aggregated choice probabilities are then

$$\pi_j = \sum_{\ell=1}^{L} \omega_\ell \pi_j(\eta_\ell;\delta,\lambda) = \sum_{\ell=1}^{L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \frac{\exp\{(\delta_j + \eta_{\ell j})/(1-\lambda_2)\}}{D_{\ell c_2}}.$$

From the equation for $\pi_{\ell c_2|c_1}$, we have that

$$D_{\ell c_2} = \left(\pi_{\ell c_2|c_1} D_{\ell c_1}\right)^{\frac{1-\lambda_1}{1-\lambda_2}}.$$

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

$$\pi_{\ell 0} = \frac{1}{1 + \sum_{c_1 \in C_1} D_{\ell c_1}^{1-\lambda_1}}.$$

Combining this with the equation for $\pi_{\ell c_1}$ implies

$$D_{\ell c_1} = \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_1}}.$$

We then have that,

$$D_{\ell c_2} = \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}}.$$

Plugging into the aggregate choice probabilities, gives

$$\begin{aligned}\pi_j &= \sum_{\ell \in L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}} \exp\left\{\frac{\delta_j + \eta_{\ell j}}{1-\lambda_2}\right\} \\ &= \exp\left\{\frac{\delta_j}{1-\lambda_2}\right\} \sum_{\ell \in L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda_2}\right\}\end{aligned}$$

Finally, taking logs we have our inversion:

$$\log(\pi_j) = \frac{\delta_j}{1-\lambda_2} + \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda_2}\right\}\right)$$

or

$$\delta_j = (1-\lambda_2)\left(\log(\pi_j) - \log\left(\sum_{\ell \in L} \omega_\ell \pi_{\ell c_1} \pi_{\ell c_2|c_1} \pi_{\ell c_2|c_1}^{\frac{1-\lambda_1}{1-\lambda_2}} \left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_2}} \exp\left\{\frac{\eta_{\ell j}}{1-\lambda_2}\right\}\right)\right)$$

■

Using similar logic, it is straightforward to derive the following result for an arbitrary number of nests, N:

Proposition 4 (Nested Logit - N Levels)**.** *For any set of* $\{\eta_\ell\}_{\ell=1}^{L}$ *the market share inversion takes the following analytic form,* $\forall j \in J$*,*

$$\begin{aligned}\delta_j =&(1-\lambda_N)\Bigg(\log\left(\pi_j\right)\\ &-\log\left(\sum_{\ell\in L}\omega_\ell\pi_{\ell c_1}\left(\sum_{n=2}^{N}\pi_{\ell c_n|c_1\cdots c_{n-1}}\pi_{\ell c_n|c_1\cdots c_{n-1}}^{\frac{1-\lambda_{n-1}}{1-\lambda_N}}\right)\left(\frac{\pi_{\ell c_1}}{\pi_{\ell 0}}\right)^{\frac{1}{1-\lambda_N}}\exp\left\{\frac{\eta_{\ell j}}{1-\lambda_N}\right\}\right)\Bigg).\end{aligned}$$

Appendix B

Appendix to Chapter 2

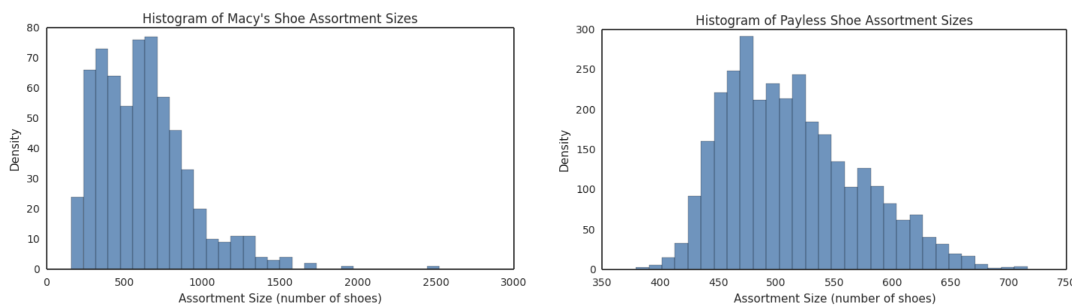
B.1 Localization in Footwear Retail

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

Table B.1: Summary of Brick-and-Mortar Data

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

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



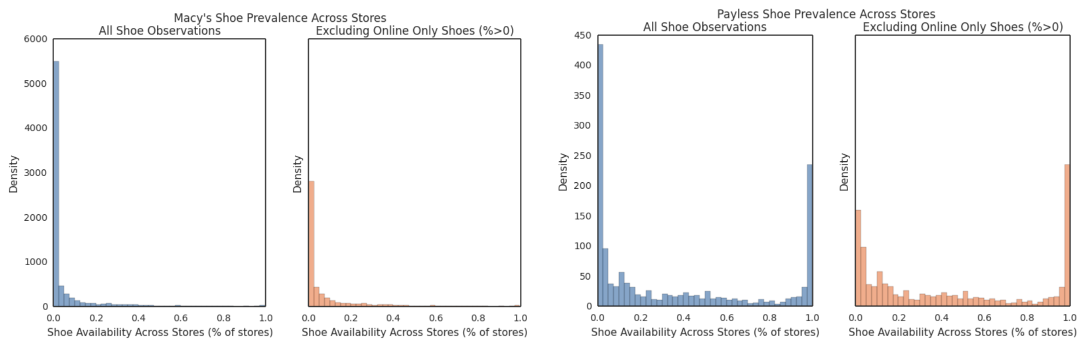
(a) Macy's

(b) Payless Shoes

Figure B.1: Shoe Assortment Size Distributions Across Retail Chains

We want to measure how assortments vary by store. Figure B.2 graphs the percentage of locations carrying a shoe style for Macy's and Payless. That is, we present a histogram of shoe presence across stores of the chain. If all shoes were available at all stores, the density would collapse at 1 (100%). The left panels within a chain plots the density for all shoes, whereas the right panel excludes online only shoes. For Macy's we can see that the vast majority of products are sold at only a few stores; that is, the density is concentrated primarily to the left. The Payless Shoes distribution is more bimodal, at a few stores and at almost all stores. In recent years, Macy's has made a concerted effort to better localize their product assortments through a program called

"My Macy's."[1] The strikingly low prevalence of products across stores is likely reflective of this program. Payless, on the other hand, produces and partners with other brands to provide exclusive products for its retail chain. The bimodal distribution for Payless may be reflective of these partnerships.



(a) Macy's (b) Payless

Figure B.2: Footwear Prevalence Across Stores

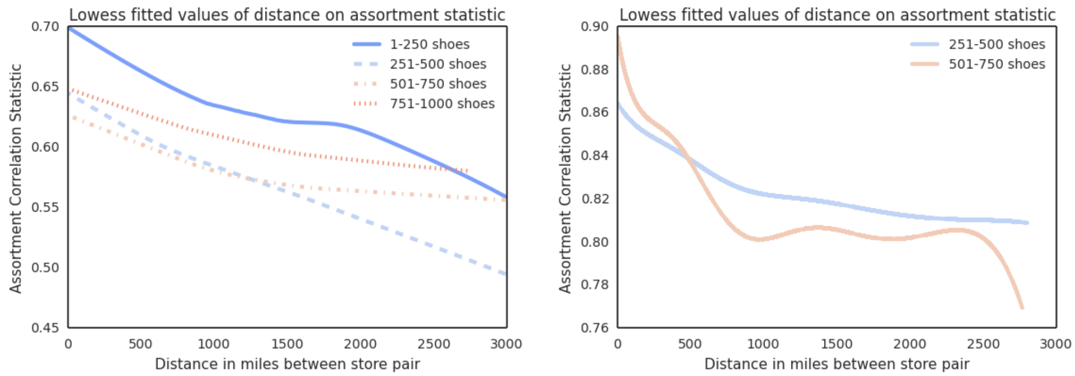
Finally, we want to measure how assortments change moving away from a particular store. To calculate this measure, we begin by taking the network of stores and create all possible links. Then for each pair of stores with assortment sets (A, B), we calculate

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

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

We see can that the assortment overlap has a decreasing relationship with distance, which suggests these retailers are localizing their product assortments. We also, note that as distance approaches zero, assortment overlap does not converge to 1. This is

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



(a) Macy's (b) Payless

Figure B.3: Assortment Overlap by Distance

likely reflective of a strategy to increase variety within a geographic area.

B.2 An Empirical Bayesian Estimator of Shares

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

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

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

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

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

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

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

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

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

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

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

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

for an unknown parameter ϑ.

Equations B.1 and B.2 then imply

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

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

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

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

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

Thus,

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

which implies

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

B.3 Additional Tables and Figures

Figure B.4: Sales Share of a Popular Brand Across Zip3s