Preference Minorities and the Internet

July 20, 2010

Preference Minorities and the Internet

Abstract

Offline retailers face trading area and shelf space constraints so they offer products tailored to the needs of the majority. Consumers whose preferences are dissimilar to the majority—*preference minorities*—are under-served offline and should be more likely to shop online. We use sales data from Diapers.com, the leading US online retailer for baby diapers, to show why geographic variation in "preference minority" status of target customers explains geographic variation in online sales. We find that, holding the *absolute* number of the target customers constant, online category sales are more than 50% higher in locations where customers suffer from preference isolation. Since customers in the preference minority face higher offline shopping costs, they are also less price-sensitive. Niche brands, relative to popular brands, show even greater online-offline sales substitution. This greater sensitivity to preference isolation means that these brands in the tail of the Long Tail distribution draw a bigger proportion of their total sales from high preference minority regions. Implications for online retailing research and practice are discussed.

Keywords: *Internet; Long Tail; Preference Minority; Retailing*

Local offline stores face trading area and space constraints, so they offer products that cater to the tastes of the local majority. Hence, when co-located consumers share preferences their individual welfare is improved as the local retail market offers products they want (Sinai and Waldfogel 2004). On the other hand, consumers whose preferences are dissimilar to the majority—*preference minorities*—are likely to be under-served, or, perhaps, neglected by local retailers altogether. In this paper, we examine how online demand for a *product category* and *individual brands* is generated from preference minorities.

We use recent findings in economics and information systems to develop our conceptual framework and hypotheses. Larger markets deliver more product variety (Glaeser, Kolko, and Saiz 2001; Waldfogel 2003) and in many ways the Internet acts like a "large market." Sinai and Waldfogel (2004, p. 3) explain: "By agglomerating consumers into larger markets, the Internet allows locally isolated persons to benefit from product variety made available elsewhere." Conceptually, two forms of isolation affect online consumer demand. The first is *geographic* isolation from offline alternatives, i.e., physical distance or transportation costs to offline alternatives. Forman, Ghose, and Goldfarb (2009), for example, show that increased distances to offline retailers leads to an increase in online demand for books. Brynjolfsson, Hu, and Raman (2009) find that better access to offline alternatives depresses online demand for apparel. The second (and our focus) is *preference* isolation—a concept that we elaborate on in the next section.

Profit-maximizing offline retailers allocate shelf space according to the Pareto or "80/20" rule (Chen et al. 1999; Reibstein and Farris 1995) and "Retail buyers favor products that provide the greatest returns to the shelf space and the merchandizing resources allotted them" (Farris, Olver, and De Kluyver 1989, p. 109). The implication is that offline retailers will pay less attention to categories favored by preference minorities, and offer smaller assortments in their stores in preference minority locations. We examine a small sample of local stores to check this assortment assumption (see Dukes, Geylani, and Srinivasan 2009, Table 1 for a similar exercise).[1] Our empirical analysis relies on sales data from Diapers.com, so we collect offline data for the diapers category. Table 1 summarizes diaper category space allocations and assortment sizes for three Fresh Grocer supermarkets and two Walmart stores in the Philadelphia area. Both chains allocate more space to diapers and carry more SKUs when the proportion of target customers (households with babies) is higher.

[Insert Table 1 about here]

More generally, in an area where the elderly are the majority of the population young parents with newborns might not find a full assortment of baby diapers at local offline retailers. That is, they assume the status of preference minorities. Local stores may still allocate *some* shelf space to baby products, but if they do, the brands and variety offered will be limited e.g., perhaps restricted to the popular packages of the leading brand, Pampers.[2] The local market characteristic of a prevalent elderly population creates preference isolation for young parents when it comes to shopping locally. If the parents have narrow *within*-category brand preferences, the effect is exacerbated (relatively limited space is allocated to the product category overall so it is even more difficult to purchase niche products locally).[3]

There is no absolute standard for defining "minor preferences" in a geographic area so we define them by looking at the *relative size* of the target group in a local area. To implement our analysis we construct a "preference minority index" (hereafter, the PM Index) for each zip code equal to [1 - (Target Population / Total Population)] (see Forman, Ghose, and Wiesenfeld 2008;

[1] We thank an anonymous *JMR* reviewer for suggesting this exercise.
[2] This preference minority effect on assortment will be especially true for products such as diapers that are bulky and have high shelf space-to-profit ratios. See Figure 2 and the related discussion.
[3] Suppose the baby is sensitive to chlorine and the parents must use chlorine-free diapers, e.g., Seventh Generation, a niche product that is on average, less readily available in local markets.

Goolsbee and Klenow 2002; Sinai and Waldfogel 2004).

Diapers.com, the largest U.S. online retailer carrying baby products, provides an excellent setting for studying geographic variation in online demand for diapers overall, and for specific brands. There are several reasons why the diapers category is well-suited for our study. First, total diaper consumption in a location is tied to the number of babies. Second, Diapers.com carries leading national brands (Pampers, Huggies, and Luvs) and a leading niche brand (Seventh Generation) that is not available in all offline retailers.[4] Third, the high shelf space-to-profit ratio for diapers limits product assortment in local markets more than for other products with lower shelf space-to-profit ratios (e.g., spices, vitamin pills, etc). Fourth, the diapers category is important to offline retailers (Kumar and Leone 1988).

We contribute four new substantive findings. First, we demonstrate that sales substitution, from offline retailers to online retailers, increases across local markets as the PM Index increases, i.e., as the *relative size* of the target group decreases (H_1). Holding the characteristics of the local environment constant, online sales are higher in markets where the target group is more of a preference minority. On average, online sales in "high PM" markets (at the 90th percentile of the PM Index) are more than 50% higher than in "low PM" markets (at the 10th percentile), *even though both these markets contain the same number of potential customers*. Second, preference isolation reduces online price sensitivity (H_2) because preference isolation implies that offline shopping costs for the category are relatively high. Model estimates imply that lowering online prices relative to offline prices increases demand by about 30% in low PM markets, but by about 10% in high PM markets. Third, local online sales of "niche" brands respond more strongly to the presence of preference minorities than local online sales of "popular" brands do (H_3). High

[4] We determine exactly which offline stores in which locations carry this brand in order to control for region-level variation in access to popular and niche brands. Details are provided in Data and Measures.

PM markets, relative to low PM markets, have local *online* sales of popular brands that are about 40% higher; yet they have local *online* sales of niche brands that are about 140% higher, *even though both markets contain the same number of potential customers*. Fourth, we find that the differential effects of preference isolation on online popular and niche brands sales have an important implication for the Long Tail sales distribution. Niche brands with a lower overall sales rank (i.e., those in the "tail" of the Long Tail) draw a greater *proportion* of their total online demand from high PM regions (H_4).

The paper is organized as follows. The next section summarizes key ideas from the literature, introduces a conceptual framework, and describes the hypotheses. The subsequent section describes the data and measures. Next, we describe the empirical model and report and interpret the findings. The paper concludes with a discussion of the implications for Internet retailing theory and practice and for future research.

CONCEPTUAL FRAMEWORK AND HYPOTHESES

Online-Offline Demand Substitution

Online retailers, relative to offline competitors, can offer consumers lower prices (Anderson et al. 2010; Brynjolfsson and Smith 2000; Goolsbee 2000), greater convenience (Balasubramanian, Konana, and Menon 2003; Forman, Ghose, and Goldfarb 2009; Keeney 1999), and more variety (Brynjolfsson, Hu, and Rahman 2009; Ghose, Smith, and Telang 2006). Among factors studied, price has received the most attention. Consumers shop online for lower prices (Brynjolfsson and Smith 2000) and to avoid local sales tax (Goolsbee 2000). Anderson et al. (2010) find that when retailers open physical stores in a location—and acquire a nexus for tax purposes—Internet sales at that location suffer (since the firm must charge sales tax on Internet orders). Forman, Ghose, and Goldfarb (2009) find that when conventional booksellers enter new offline locations Amazon.com sales at those locations decline. The increased convenience of offline alternatives reduces the attractiveness of the online alternative.

Finally, Brynjolfsson, Hu, and Rahman (2009) show that a consumer living in an area with the median number of U.S. apparel stores nearby has Internet demand that is 4.2% lower than another consumer with no offline stores nearby. They conclude that "Internet retailers face significant competition from brick-and-mortar retailers when selling mainstream products, but are virtually immune from competition when selling niche products" (p. 1755). The focal variable in this study is a measure of offline "search and transaction costs"; specifically, the number of offline stores nearby. Here, we provide additional insight by considering the preference isolation of local customers. It reduces the Internet retailer's competition for niche products and popular products, i.e., it is *demand enhancing* for the Internet retailer, for both types of products.

Preference Isolation and Preference Minorities

Co-location of several consumers with shared needs produces two effects in a trading area. First, offline retailers pay more attention to the product category this group wants. Second, individual consumers are more able to find and buy products that suit their needs locally. This effect on local assortment is especially evident when the fixed cost of product provision is high. Media markets, for example, have high fixed costs so "specialty products" like Spanish programs emerge only when sufficient numbers of customers demand them (Waldfogel 2003). Similarly, shelf space constraints make this fixed cost argument relevant to offline retailers. Table 1 provided some preliminary evidence that shelf space allocations and assortment sizes decrease when the target customer group becomes “less important”, i.e., makes up a smaller proportion of the total market. Figure 1 conceptualizes the key relationships between preference isolation, store shelf space decisions, and the resulting assortments. It presents two hypothetical markets to show how *geographic variation* in preference isolation will affect online demand for a product category. Both markets have the same number of consumers in the target population (100) however the target group makes up 50% of the consumers in Market A, but only 10% in Market B. Since Market B is larger, it contains more stores. It is well known (and perhaps obvious) that larger markets have more stores overall (see Christaller 1933 for an explanation of “Central Place Theory”, a descriptive view of how the number of retail stores grows with population size).

[Insert Figure 1 about here]

Stores in both markets allocate category space in line with the size of the target population (Chen et al. 1999). Each *individual store* in Market B pays less attention to the target group (allocating only 10% of each store to the category), yet both markets devote the same *total amount of space* to the product category.[5] Critically, *store*-level shelf space allocation to categories based on the relative size of the target population produces more offline assortment in Market A compared to Market B, even though the size of the target customer group and the *aggregate* shelf space allocated to the category are identical in both markets. In offline stores leading brands, e.g., Pampers, tend to be stocked in all stores whereas “niche” brands are stocked only in stores with considerable shelf space (Farris, Olver, and De Kluyver 1989; Reibstein and Farris 1995).

Before using this conceptual framework to develop our hypotheses, it makes sense to validate

[5] It is reasonable to assume that since both markets have the same aggregate need for the product category, both will devote an equal amount of attention, in aggregate. We thank the AE for this clarifying observation.

the three key assumptions in Figure 1 using prior research findings and market data from the 2007 US Census of Business and Industry.

- *Assumption 1*: *Shelf Space Allocations*. Prior work by Chen et al. (1999) rests on the assumption that space allocations are proportional to the size of the customer group. This is also suggested by Table 1.
- *Assumption 2*: *Total Population and Stores*. The correlation between total population and the number of local stores is strongly positive. At the Metropolitan Statistical Area (MSA) level the correlations are .97, .86, and .96, for supermarkets, discount stores (Walmart and Target), and warehouse clubs, respectively.
- *Assumption 3*: *Constant Store Size.* Store size tends to be uniform within a retail chain. Among the 1,415 Target stores in our data, for example, 99% belong to the highest size range (> 40,000 sq ft) and 81% belong to a single employee number range (100-249 employees).[6]

In summary, it is reasonable to assume that retailers allocate space according to the size of the target group and that while the number of stores increases with population, the size of individual stores in a given chain does not. Moreover, stores with less space devoted to a category focus on popular brands, so many niche brands do not "make the cut" (Anderson 1979). Hence, the amount of local product variety available *offline* to the target group depends on the *relative* size of the target group.

Figure 2 shows some preliminary evidence for a positive relationship between preference isolation and online sales. Recall that the PM Index is equal to [1 - (Target Population / Total Population)]. Figure 2 (a) maps the five quintiles of the PM Index in Los Angeles County and Figure 2 (b) maps the cumulative number of orders per target household placed at Diapers.com for the 39 months of our data. Shading patterns show a positive correlation: in zip codes where households with babies are in the minority, online sales per target household are higher.

[Insert Figure 2 about here]

[6] For retail stores the 2007 US Census of Business and Industry uses 4 physical size (sq ft) bins and 11 employee number bins.

Hypotheses

Using the ideas developed in Figures 1 and 2 we specify four different effects of preference isolation on geographic variation in online demand. The first two hypotheses are for overall product category effects; the next two specify differential effects for popular and niche brands.

Category Sales Online (H_1). Offline retailers' assortment decisions react to preference isolation and affect the amount of total local demand satisfied online versus offline. When a product category has fixed per-capita consumption (conditional on controlling for geographic variation in other important covariates such as income and so forth), so that markets with the *same number* of target customers have the *same* aggregate demand, preference isolation predicts geographic variation in the amount of demand satisfied online.

H_1: Substitution from offline retailers to online retailers will be greater in markets that have a higher PM Index.

Category Price Sensitivity Online (H_2). It is not possible to measure offline prices of Diapers.com competitors, however prior research suggests that at the *geographic market level* offline tax rates reflect the *relative advantage* of online prices over offline prices for homogenous goods (Anderson et al. 2010; Goolsbee 2000). Hence, geographic variation in the relative advantage of Diapers.com prices can be captured by geographic variation in offline sales taxes.

At the individual product level, Brynjolfsson and Smith (2000) find that for homogenous goods (books and CDs) online prices are lower than offline prices by 9-16% implying that online shoppers might be more price-sensitive than their offline counterparts or, at the very least, lower prices are a reason to shop online. At the same time Lynch and Ariely (2000) show that online shoppers can be made less price sensitive and more quality sensitive, depending on how category information is presented. Like Brynjolfsson and Smith (2000) we study a homogenous good with known *ex ante* quality (Lynch and Ariely 2000 use wine so product quality is ambiguous for many consumers).

Preference isolation suggests additional geographic variation in price-response, over and above the direct positive effect of the online price advantage on online demand. Preference isolation increases category shopping costs in the offline market; products are less accessible, and the category is less well assorted. This increases the relative value of the online alternative, holding the price advantage constant.

H_2: Markets with a higher PM Index show lessened sensitivity to the online price advantage.

Brand Sales Online: *Popular versus Niche* (H_3). Offline retailers prioritize shelf space in a category. All else equal, they stock popular brands such as the leading national brand before adding niche brands to their assortments (e.g., Farris, Olver, and De Kluyver 1989; Reibstein and Farris 1995). Preference isolation creates double jeopardy for niche brands; fewer consumers prefer them to begin with, and in high preference minority markets retailers pay *even less* attention to these brands. Category-level online-offline substitution predicted in H_1 is intensified for niche brands.

H_3: Online-offline substitution for niche brands, relative to popular brands, is *more* sensitive to geographic variation in the PM Index.

Brand Sales Online and the Long Tail (H_4). Anderson (2006) popularizes the Long Tail sales distribution concept; namely, the idea that the Internet allows sellers to stock more variety and relatively low sales for individual niche brands combine to contribute up to 20% of total sales and profits. Products that would not "make the cut" at offline retailers can be profitably sold by online retailers (Brynjolfsson, Hu, and Smith 2006). H_3 predicts that online sales of niche brands also respond strongly to preference isolation. If H_3 holds, it follows immediately that the sales decomposition over space will differ for popular and niche brands.

H_4: Niche brands with a lower overall sales rank, i.e., those in the "tail" of the Long Tail, draw a greater *proportion* of their total online demand from high PM regions, than popular brands do.

DATA AND MEASURES

Here we describe our online sales data, reiterate why the diapers category is well-suited to our research, and define the unit of analysis for a "local market". We also describe the variables that control for geographic differences across offline markets.

Product Category and Unit of Analysis

Product Category. Diapers.com, the leading US online retailer for baby diapers provided: (1) zip-level cumulative numbers of buyers and orders from the website's inception in January 2005 through March 2008, and (2) zip-level cumulative sales by brand between January 2007 and March 2008. We use the three major national brands—Pampers, Huggies, Luvs—and one niche brand that is not available in all stores, Seventh Generation, in our analysis.[7] Summary statistics for the dependent variables are presented in Table 2 (a). Orders over $49 (about 90% of all

[7] Seventh Generation limits distribution to bricks-and-mortar retailers that have an image of being "natural" or "organic" (e.g., Whole Foods). We control for these store locations in the empirical analysis.

orders) qualify for free shipping and are shipped via UPS from Diapers.com warehouses. Diapers.com made no marketing interventions or promotional efforts targeted at preference minority regions or customers; hence, we can assess how preference isolation in a region affects online demand there, free from explicit marketing interventions.

[Insert Table 2 about here]

As noted in the Introduction, diapers are especially suitable for our study. In addition to reasons given previously, the fact that the brands are well known and shoppers face little (if any) quality uncertainty means the products have no "non-digital attributes" (Lal and Sarvary 1999). The fact that Diapers.com is a *category*-focused online retailer is also ideal because in our study preference isolation is a category-level phenomenon (see Figure 1).

Unit of Analysis. The zip code is the unit of analysis; this makes sense for two reasons. First, zip codes encompass relatively self-contained groups of buyers and sellers for packaged goods such as diapers. The most accessible offline local retail format for diapers is the local supermarket and all zip codes that we examine have at least one supermarket.[8] (Of course we also control for other offline retail alternatives.) Second, zip codes are used in many related studies of retail phenomena (see Waldfogel 2007 for a review) and following this literature, we focus on zip codes within Metropolitan Statistical Areas (MSAs).[9] Limiting the analysis to zip codes within MSAs ensures shoppers have "reasonable" travel distances to offline alternatives, i.e., they don't shop online due to a complete lack of offline stores, and is also consistent with

---

[8] Residential zip codes have on average four supermarkets. There is roughly one discount store for every five zip codes, and one warehouse club for every fifteen zip codes. Supermarkets appear at approximately 2.5 miles intervals, discount stores every 8 miles, and warehouse clubs every 15 miles.

[9] MSAs are formed around a central urbanized area with surrounding areas that have "strong ties" to the central area. This spatial demarcation is more comprehensive than one based on just geographical boundaries. (Delaware Valley is an MSA comprising counties in DE, NJ, MD, and PA. There are 358 MSAs in the 48 contiguous states.)

prior research (e.g., Brynjolfsson, Hu, and Raman 2009; Forman, Ghose, and Goldfarb 2009; Sinai and Waldfogel 2004).

Geographic Variation in Online Shopping Costs, Market Potential, and Demographics

Hypotheses H_1-H_4 predict geographic variation in online demand as a function of geographic variation in preference isolation. Hence, we need to control for geographic variation in overall online shopping costs, and other known demographic factors that affect the propensity for shoppers to buy online. Online shopping costs consist of price, waiting time, and convenience costs. Descriptive statistics for all independent variables are given in Table 2 (b).

Price. Diapers.com offers the *same* product prices in every zip code but shoppers in different zip codes face different *offline* product prices. It is not possible to gather offline prices for diapers in every US zip code, however prior research (Anderson et al. 2010; Goolsbee 2000) shows that *geographic* variation in offline prices for homogenous goods can be captured using geographic variation in offline sales tax for those goods. In zip codes where offline stores collect sales tax on diapers, Diapers.com has a greater *relative price advantage* over offline competitors, compared to zip codes where offline taxes are not collected.[10] When offline stores collect sales tax on diapers, online shoppers have lower shopping costs so higher offline sales taxes should lead to higher online demand. We use data from the Department of Revenue in each state to determine offline sales taxes for diapers.[11]

[10] The presence of the lowest-priced offline competitor (Walmart and warehouse clubs) creates lower offline diaper prices in a zip code so we run additional models with dummy variables for these stores. These dummies and their interactions with the PM Index are not significant (details are available upon request). We thank an anonymous reviewer for this suggestion. Insignificant effects could also due to Diapers.com striving for “Walmart-level pricing” on diapers (communication with management).

[11] Our offline diapers tax data are very detailed. We made over 1,000 telephone calls to a random sample of major offline retailers and asked store employees to determine whether diapers were tax exempt.

Waiting Time. Shipping times reflect online shopping convenience (Brynjolfsson and Smith 2000); hence we need to control for geographic variation in shipping times. Shoppers are informed of the shipping time to their zip code when they order and we collect zip-level shipping time data from UPS. Shorter waiting times reduce online shopping costs and so zip codes receiving faster shipping should see more online demand.

Convenience. The attractiveness of shopping online is affected by the availability of offline stores (Brynjolfsson, Hu, and Rahman 2009) and the travel distance to offline stores (e.g., Forman, Ghose, and Goldfarb 2009). We use 8-digit NAICS (North American Industry Classification System) codes from the 2007 US Census of Business and Industry to measure the convenience of relevant offline stores.[12] Physical distance reflects transportation costs in spatial differentiation models (see e.g., Balasubramanian 1998; Bhatnagar and Ratchford 2004) so we use the actual store locations of all supermarkets, discount stores (Walmart and Target) and warehouse clubs in the database and compute the expected distance to each type of store for residents of each zip code.[13] Finally, Pampers, Huggies, and Luvs have extensive offline distribution but Seventh Generation diapers do not. Hence, we collect location data for all stores carrying Seventh Generation and separately compute the distance from each zip code to the nearest store where this brand is available (in order to test H_3). Greater distances to offline retailers make online shopping more convenient and should lead to higher online demand.

Market Potential and Demographics. Zip codes with greater potential for the diaper category should have more specialist retailers targeted at households with babies. To control for this we count the number of stores selling baby accessories, e.g., Babies R Us, using NAICS 44813001

[12] Prior research sometimes uses 6-digit NAICS codes but our use of 8-digit codes, while more laborious, leads to greater accuracy in store classification. As an example, using 6-digit codes can lead "candy stores" to be included with "supermarkets" but using 8-digit codes does not.
[13] We also re-estimate the models controlling for distance to the *second* store of each format and obtain qualitatively identical results (available upon request). We thank an anonymous reviewer for suggesting this robustness check.

and 44813002. Other relevant control variables are created from 2000 US Census of People and Households. Measures of income and education control for the propensity to shop online and opportunity costs of time, while age, target population, and population density help control for overall market potential.

Preference Isolation and the PM Index

As noted earlier, there is no absolute determinant for "minority preferences" so we focus on the proportion of households with babies aged less than 6 years old and define the PM Index at the zip code level as: [1 - (Households with Babies / Total Households)]. As argued in the Conceptual Framework and Hypotheses section, geographic variation in this measure will reflect geographic variation in offline retail assortments offered to households with babies. Diapers.com, on the other hand, offers the same assortment in all zip codes.

MODEL AND EMPIRICAL FINDINGS

We now describe the testing approach and the empirical findings for H_1-H_4. We also report robustness checks which include an additional check of the overall "process" argument, i.e., that preference isolation is negatively correlated with offline variety in a zip code (see also Table 1).

Preference Isolation and Category Sales Online

We test H_1 and H_2 with two dependent variables: (1) the number of buyers per zip code, and (2) the number of repeat orders per zip code. We assume that the number of buyers (repeat orders) in zip code *z* in MSA *m* is Poisson distributed with rate parameter $\lambda_{z(m)}$ .The Poisson is appropriate when the rate of occurrence is low (Agresti 2002) and in our data the number of buyers and repeat orders in a zip code is small relative to the number of households with babies. The number of households with babies aged less than 6 years old, $n_{z(m)}$, serves as an offset variable with its parameter constrained to one (Knorr-Held and Besag 1998; Michener and Tighe 1992).

The Poisson rate in zip code z and MSA m, $\lambda_{z(m)}$, is modeled as a function of the PM Index, $PM_{z(m)}$, zip code offline sales tax rate, $TAX_{z(m)}$, the interaction of these two variables, and the other shopping cost and control variables discussed previously in Data and Measures.

(1) $$y_{z(m)} \sim \text{Poisson}(\lambda_{z(m)}) \text{ and}$$

(2) $$\begin{aligned} \log(\lambda_{z(m)}) &= \beta' x_{z(m)} + \varepsilon_{z(m)} \\ &= \beta_1 \cdot PM_{z(m)} + \beta_2 \cdot TAX_{z(m)} + \beta_3 \cdot PM_{z(m)} \cdot TAX_{z(m)} \\ &\quad + \log(n_{z(m)}) + \vec{\gamma}' \cdot Controls_{z(m)} + \alpha_0 + \alpha_m + \varepsilon_{z(m)} \end{aligned}$$

$$\text{where } \alpha_m \sim N(0,\tau^2) \text{ and } \exp(\varepsilon_{z(m)}) \sim Gamma(\theta,\theta).$$

The baseline rate for regional cluster m consists of the overall baseline, α_0, and the deviation of MSA m from the overall baseline, α_m. These MSA-level random effects control for unobserved heterogeneity in the baseline rates.[14] The error tem $\varepsilon_{z(m)}$ allows for over-dispersion and is *IID* Gamma distributed with shape and scale parameters both equal to θ for identification (Cameron and Trivedi 1986; Greene 2008). After integrating over $\varepsilon_{z(m)}$ the density for $y_{z(m)}$ becomes one form of the negative binomial distribution with mean $\mu_{z(m)}$ and variance

[14] We estimate equation (2) with MSA-level fixed effects and find qualitatively identical results, e.g., for the number of buyers β_1 = 4.415 versus 4.479 in Table 3 (details available upon request). The Hausman test suggests that a fixed effects model is preferred; however, given the nearly identical estimates for the category demand model we report random effects results in Table 3. This is also for consistency with our brand demand models that use a multivariate NBD model (Gueorguieva 2001; Thum 1997) with multivariate random effects (see equation 5) to parsimoniously accommodate correlations in zip-level brand demands. We thank an anonymous reviewer for suggesting the fixed effect check.

$\mu_{z(m)}\left(1+\theta^{-1}\mu_{z(m)}\right)$.[15] Our model has a closed-form solution up to the MSA-level random effects and the likelihood is evaluated via numerical integration over the random effects.

Category Sales Online (H_1). A larger PM Index means that the target customer group suffers from more preference isolation, which should increase the attractiveness of buying online, i.e., we expect $\beta_1 > 0$. Table 3 shows that $\beta_1 = 4.479$ ($p < .001$) for the number of buyers and $\beta_1 = 5.644$ ($p < .001$) for the number of orders. These estimates imply economically meaningful effects which can be seen by comparing zip codes at different deciles on the PM Index (see Sinai and Waldfogel 2004 for a similar analysis). This standard computation of marginal effects also assumes that the zip code location of the household is exogenous to the decision to use Diapers.com (see also Forman, Goldfarb, and Greenstein 2005, p. 398). Suppose that two zip codes have the same number of shoppers in the *target* population, but differ in terms of *total* population, and therefore on the PM Index. Suppose one compares online category demand in a "low PM market" (the 10^{th} percentile market; PM Index = .79) and a "high PM market" (the 90^{th} percentile; PM Index = .89).[16] At the mean of the other covariates, this implies 6.66 buyers and 9.86 repeat orders in the low PM market but 9.86 buyers and 16.31 repeat orders in the high PM market. Taking trial and repeat orders together, this implies that Diapers.com sales are about 57% higher in the high PM market *even though both markets have the same number of target consumers*. Thus, H_1 is strongly supported.

[Insert Table 3 about here]

[15] Brynjolfsson, Hu, and Rahman (2009) also use the NBD model to study region-level variation in online demand.

[16] H_1 says that the slope coefficient for the PM Index should be positive, not that the slope will increase as the PM Index increases. To check this we allow for heterogeneity in the coefficient of the PM Index across the four quartiles of its value in a NBD spline regression. The coefficient for the PM Index is essentially constant ranging from 4.346 to 4.499. We thank an anonymous reviewer for suggesting this check.

Category Price Sensitivity Online (H_2). A higher offline tax rate means shoppers in the zip code have relatively lower online shopping costs, which should increase the attractiveness of buying online, i.e., we expect $\beta_2 > 0$. Table 3 shows that $\beta_2 = .113$ ($p = .003$) for the number of buyers and $\beta_2 = .164$ ($p = .028$) for the number of orders. H_2, however, is not about this straightforward main effect, but instead about the interaction between preference isolation and price sensitivity. Shoppers suffering from preference isolation face higher shopping costs; hence, they need less price-based inducement to shop online. Since the Poisson/NBD model has a non-linear functional form, we evaluate the interaction effect by computing the cross derivative and applying the Delta method (Ai and Norton 2003), rather than simply looking at the significant interaction parameters only.[17]

The parameter estimates in Table 3 show the expected results for the number of buyers and repeat orders. Using the model estimates, we compute expected online demand by varying both the PM index and the offline sales tax rate (which ranges from zero to 8.25%). When online shopping costs are lower, i.e., offline tax rates are higher, Diapers.com demand in low PM markets increases by about 27%. This same reduction in shopping costs also increases Diapers.com demand in high PM markets, but only by 12%. Thus, we find strong support for H_2 as well.

Control Variables. Control variables for online shopping costs and demographics were selected according to prior studies and generally have expected signs. As noted earlier, higher offline tax rates reduce online shopping costs and increase demand. Less waiting time, e.g., 1-day shipping versus 2-day shipping reduces online shopping costs and increases online demand.

[17] In a non-linear model such as the Poisson/NBD the sign of an interaction term is not necessarily the same as the sign of a marginal effect (Ai and Norton 2003). In fact, Ai and Norton (2003, p. 123) report a disturbing finding: "A review of the 13 economics journals listed on JSTOR found 72 articles published between 1980 and 1999 that used interaction terms in nonlinear models. None of the studies interpreted the coefficient on the interaction term correctly." We thank an anonymous reviewer for providing this reference.

Larger distances to offline stores should reduce online shopping costs and increase online demand. This intuitive effect is present and significant for discount stores (Walmart and Target) and warehouse clubs.[18] Overall market potential, as indicated by the presence of “baby-oriented” stores has a weaker but positive effect on online demand.

Online demand also increases with known demographic drivers. Not surprisingly, Diapers.com performs better in zip codes that have higher percentages of people between 20 and 39 years old, more working females, more urban housing units, and more homes valued in excess of $250,000 or more. Less online demand in areas with higher percentages of black households and households below the poverty line is also consistent with the commonly held notion of a “digital divide” (e.g., digitaldivide.org). Finally, zip codes with greater population density and population growth have higher online demand.[19]

One less straightforward finding is the negative and significant coefficient for distance to supermarkets (although this effect on online demand was also observed by Bell and Song 2007 for a different online retailer). Greater distances to supermarkets might be expected to reduce online shopping costs, and therefore increase online demand. A possible explanation for opposite conclusion for supermarkets implied by the negative coefficient is that most shoppers will visit a supermarket regardless of whether they buy diapers online. Shoppers who have to travel further to offline stores might try to amortize fixed shopping costs by buying larger baskets of items per trip (Tang, Bell, and Ho 2001), which reduces their need for an online retailer.

In summary, it is reassuring to see that the hypothesized preference isolation effects hold in

[18] Note that our findings on the effect of preference isolation are obtained in a model that also controls for “geographical isolation” (transportation costs) studied in other articles (e.g., Brynjolfsson, Hu, and Rahman 2009; Forman, Ghose, and Goldfarb 2009).
[19] Growing areas are more likely to have new residential buildings with young families while local retailers are less available. This negative correlation between population growth and the PM Index (r = -.254 in our data) could inflate the coefficient for the PM Index. We re-estimate the models without the population growth variable and obtain qualitatively identical results (β_1 = 4.220, p < .001for the number of buyers; further details available upon request). We thank an anonymous *JMR* reviewer for suggesting this check.

the presence of significant effects for other well-established control variables which account for geographic variation in online shopping costs and demographics.

Alternative Process Evidence

Our conceptual framework is built on the premise that preference isolation affects online sales by influencing the extent of product variety offered by offline retailers. We provide a more direct test to shed some light on the process. The logic here is that first, niche brands will be less available in high PM markets and second, there will be greater online demand when niche brands are less available.[20] First, we estimate a probit model of the zip-level presence of local stores selling Seventh Generation diapers as a function of the same set of local variables in Equation (2). The significantly negative effect of the PM Index (estimate = -17.232, $p < .001$) implies that local stores selling Seventh Generation diapers are less available in high PM markets. Second, we estimate the NBD model of the zip-level online demand as a function of the same set of local variables in Equation (2) but after replacing the PM index with the dummy variable for the presence of local stores selling Seventh Generation diapers. The significantly negative effect of the store dummy variable (estimate = -.046, $p = .012$) implies that online sales are lower in local markets with retailers carrying Seventh Generation diapers, i.e., markets with more offline variety. This two stage empirical examination on the "process" provides further support for our underlying conceptual framework (e.g., Figure 1).

Preference Isolation and Brand Sales Online

[20] We thank an anonymous *JMR* reviewer for suggesting both the idea and the form of this process check.

To test H_3, we estimate a multivariate model where the dependent variable, $y_{i,z(m)}$, is the number of brand *i* diaper "standard packages" purchased in each zip code *z* and MSA *m* (*i* = Pampers, Huggies, Luvs, and Seventh Generation) and follows a Poisson distribution.[21] As before, the Poisson rate, $\lambda_{i,z(m)}$, is modeled as a function of the PM Index, $PM_{z(m)}$, zip code offline sales tax rate, $TAX_{z(m)}$, the interaction of these two variables, and the other shopping cost and control variables discussed previously in Data and Measures. The number of households with babies, $n_{z(m)}$, again serves as an offset and the marginal density of $y_{z(m)}$ becomes one form of the negative binomial distribution.

(3) $$y_{i,z(m)} \sim \text{Poisson}(\lambda_{i,z(m)}) \text{ and}$$

(4) $$\begin{aligned}\log(\lambda_{i,z(m)}) &= \beta_i' x_{z(m)} + \varepsilon_{i,z(m)} \\ &= \beta_{i,1} \cdot PM_{z(m)} + \beta_{i,2} \cdot TAX_{z(m)} + \beta_{i,3} \cdot PM_{z(m)} \cdot TAX_{z(m)} \\ &\quad + \log(n_{z(m)}) + \vec{\gamma}_i' \cdot Controls_{z(m)} + \alpha_{i,0} + \alpha_{i,m} + \varepsilon_{i,z(m)}\end{aligned}$$

$$\text{where } \alpha_{i,m} \sim N(0,\tau_i^2) \text{ and } \exp(\varepsilon_{i,z(m)}) \sim Gamma\left(\theta_i, \theta_i\right)$$

Since online demand from each of the four brands emerges from the same regional cluster we include four random effects that follow a multivariate normal distribution (Gueorguieva 2001; Thum 1997). Joint estimation with a single model allows us to compare the effect of one covariate, e.g., the PM Index, across brands.

(5) $$\begin{pmatrix} \alpha_{Pampers,m} \\ \alpha_{Huggies,m} \\ \alpha_{Luvs,m} \\ \alpha_{SeventhGeneration,m} \end{pmatrix} \sim \text{i.i.d.}\, MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & r_{12}\tau_1\tau_2 & r_{13}\tau_1\tau_3 & r_{14}\tau_1\tau_4 \\ r_{12}\tau_1\tau_2 & \tau_2^2 & r_{23}\tau_2\tau_3 & r_{24}\tau_2\tau_4 \\ r_{13}\tau_1\tau_2 & r_{23}\tau_2\tau_3 & \tau_3^2 & r_{34}\tau_3\tau_4 \\ r_{14}\tau_1\tau_4 & r_{24}\tau_2\tau_4 & r_{34}\tau_3\tau_4 & \tau_4^2 \end{pmatrix} \right)$$

---

[21] Each diaper SKU has a different number of actual diapers. We standardize across SKUs by converting all counts to standard unit based on the most frequently purchased package sizes. The SKU "Pampers Swaddlers Jumbo Pack Size 2 – 80 counts", for example, converts to 2 packages of "Pampers Swaddlers Super Mega Pack Size 2 – 40 counts." This approach to "standard units" mirrors the way SKUs with multiple sizes are treated in scanner panel data (see, for example, Bucklin, Gupta, and Siddarth 1998).

The model has a closed-form solution up to the regional random effects and the likelihood is evaluated via numerical integration over the multivariate random effects; however, evaluation demands increase with the dimension of the random effects vector. We alleviate this by fitting all pair-wise bivariate models separately and calculating the estimates and their sampling variation for the full multivariate model (Fieuws and Verbeke 2006; Fieuws et al. 2006).

Brand Sales Online: Popular versus Niche (H_3). A larger PM Index means that the target customer group suffers from more preference isolation. This not only makes it more attractive to buy the category online but also makes it *especially* attractive to buy niche brands online, as they suffer the most when category shelf space is reduced (Farris, Olver, and De Kluyver 1989). We expect $\beta_1 > 0$ for all brands and that β_1 for Seventh Generation is greater than those for the national brands (which are not different from each other). Table 4 shows that for Seventh Generation $\beta_1 = 7.741$ ($p < .001$), that this estimate is larger than the corresponding estimates for the national brands, and that these national brand estimates are not different from each other. Thus, H_3 is supported.

The implied quantitative effects show that online sales of the niche brand benefit disproportionally from preference isolation. At the mean of the other covariates a high PM market, compared to a low PM market, generates about 40% more online demand for the leading brand (Pampers). The increase for the niche brand (Seventh Generation) is dramatically greater at almost 140%, albeit from a smaller sales base level of sales.

[Insert Table 4 about here]

Brand Sales Online and the Long Tail (H_4). Online sales of niche brands show the strongest response to preference isolation (H_3) and this immediately implies that niche brands will draw a greater *proportion* of their total online demand from high PM markets.[22] Figure 3 (a) is a Long Tail plot of expected sales for the four brands (*x*-axis = brands ranked by expected sales, *y*-axis = sales). The three shaded bars compute expected sales from high, median, and low PM markets. Figure 3 (b) shows the percentage decomposition. The national brands have very similar decompositions and draw roughly 1.2 to 1.4 times more sales from high PM markets compared to low PM markets. The decomposition for the niche brand (Seventh Generation) shows a stark contrast. The ratio of sales from the high to low PM markets is 48:20, or about 2.5:1. Preference isolation is especially conducive to online sales of niche brands in line with H_4.

[Insert Figure 3 about here]

CONCLUSION

Drawing on theory and empirical findings in economics and information systems, we introduce the concept of preference isolation as a driver of online-offline sales substitution in local markets. We find that category-level sales substitution to online retailers is greater in markets that have a higher PM Index (H_1), high PM markets are less price sensitive (H_2), online-offline substitution due to preference isolation is significantly greater for niche brands (H_3) and that niche brands therefore draw a greater *proportion* of their total sales from high PM markets (H_4).

Implications for Online Retailing

[22] This also implies that high PM markets have more heterogeneous online demand across brands than low PM markets do (online demand in low PM markets is more concentrated on the leading popular brand while high PM markets have a more “even” distribution of online purchased assortments). This is noteworthy because it says that across-market diversity in online brand choices can be explained by the preference isolation that consumers face in their local markets, rather than simply preference heterogeneity.

Our findings imply an important geographic targeting heuristic for Internet retailers. It is natural to focus on markets where the *absolute* number of potential customers is high however customers in these markets will be well served by offline retailers. Internet retailers must also consider the *relative size* of their target customer group in a given location. Overall online category sales in high PM markets, relative to low PM markets, can be more than 50% higher, even though both markets have the same total number of customers who need the category. Selling niche brands in high PM markets is especially attractive because customers there face high offline shopping costs and are less price-sensitive. Offline retailers can improve the economics of stocking slower-moving SKUs (and therefore increase the variety they offer) by using distributors who stock in less-than-case pack-out quantities. Even so, several categories (e.g., bulky or low value) or brands (e.g., niche) that need minimum facing and are difficult for offline retailers to justify. Online retailers can exploit this assortment gap, particularly in high PM markets. Although we demonstrate this for the diapers category other categories with similar properties should also benefit in the same way.

Our study adds to the body of evidence that consumer benefits from online shopping are contextual and need to be assessed relative to offline options (Anderson et al. 2010; Brynjolfsson, Hu, and Raman 2009; Choi, Hui, and Bell 2010; Forman, Ghose, and Goldfarb 2009). In particular, we show how a specific form of consumer isolation—preference isolation—explains geographic variation in online demand. The net benefit of online shopping for individual consumers depends not only on where they live but also on *who* lives next to them.

Limitations and Future Research

First, we bring the concept of preference isolation to bear on the substantive marketing issues of assortment and online retail demand. Further opportunities exist for theory development in how Internet-facilitated connectivity affects behavior. Agrawal and Golfarb (2008) show that the Bitnet reduced "isolation" facilitated an approximately 40% increase in multi-institutional research collaboration among engineering faculty. Similarly, Internet retailer consumers might shape assortment offers and otherwise benefit from consumer-to-consumer interaction.

Second, we analyze preference minorities and develop the preference isolation concept from a cross-sectional perspective. Given appropriate data one could examine the evolution of preference minority status over time, and perhaps explore the dynamic nature of substitution between online and offline markets (e.g., Overby and Jap 2009). The possibility of endogenous preference for variety could also be examined—preference minorities might go online for the reasons we suggest (H_1), but having got there, expand their brand preferences within a category. (In our data preferences appear stable as there is little switching among brands, conditional upon shopping at Diapers.com). We intend to pursue these issues in future research.

REFERENCES

Agrawal, Ajay and Avi Goldfarb (2008), "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 98 (4), 1578-1590.

Ai, Chunrong and Edward C. Norton (2003), "*Interaction Terms in Logit and Probit Models*," *Economic Letters*, 80, 123-129.

Anderson, Chris (2006), *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion: New York, NY.

Anderson, Eric T., Nathan M. Fong, Duncan I. Simester, and Catherine E. Tucker (2010), "How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet," *Journal of Marketing Research*, 47 (2), 229-239.

Anderson, Evan E. (1979), "An Analysis of Retail Display Space: Theory and Methods," *Journal of Business*, 52 (1), 103-118.

Agresti, Alan (2002), *Categorical Data Analysis*, Wiley: New York, NY.

Balasubramanian, Sridhar (1998), "Mail versus Mall: A Strategic Analysis of Competition between Direct Marketers and Conventional Retailers," *Marketing Science*, 17 (3), 181-195.

________, Prabhudev Konana, and Nirup M. Menon (2003), "Customer Satisfaction in Virtual Environments: A Study of Online Investing," *Management Science*, 49 (7), 871-889.

Bell, David R. and Sangyoung Song (2007), "Neighborhood Effects and Trial on the Internet: Evidence from Online Retailing," *Quantitative Marketing and Economics*, 5 (4), 361-400.

Bhatnagar, Amit and Brian T. Ratchford (2004), “A Model of Retail Format Competition for Non-Durable Goods,” *International Journal of Research in Marketing*, 21 (1), 39-59.

Brynjolfsson, Erik and Michael D. Smith (2000), “Frictionless Commerce? A Comparison of Internet and Conventional Retailers,” *Management Science*, 46 (4), 563–585.

________, Yu (Jeffrey) Hu, and Michael D. Smith (2006), “From Niches to Riches: Anatomy of the Long Tail,” *MIT Sloan Management Review*, 47 (4), 67-71.

________, ________, and Mohammad S. Rahman (2009), “Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition,” *Management Science*, 55 (11), 1755-1765.

Bucklin, Randolph E., Sunil Gupta, and S. Siddarth (1998), “Determining Segmentation in Sales Response across Consumer Purchase Behaviors,” *Journal of Marketing Research*, 35 (2), 189-197.

Cameron, A. Colin and Pravin K. Trivedi (1986), “Econometric Models Based on Count Data” Comparisons and Applications of Some Estimators and Tests,” *Journal of Applied Econometrics*, 1, 29-54.

Chen, Yuxin, James D. Hess, Ronald T. Wilcox, and Z. John Zhang (1999), “Accounting Profits Versus Marketing Profits: A Relevant Metric for Category Management,” *Marketing Science*, 18 (3), 208-229.

Choi, Jeonghye, Sam K. Hui, and David R. Bell (2010), “Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer,” *Journal of Marketing Research*, 47 (1), 75-89.

Christaller, Walter (1933), *Die zentralen Orte in Suddeutschland*, Jena: Gustav Fischer. (Translated (in part) by Charlisle W. Baskin (1966), as *Central Places in Southern Germany*, Englewood Cliffs, NJ: Prentice Hall).

Dukes, Anthony J., Tansev Geylani, and Kannan Srinivasan (2009), "Strategic Assortment Reduction by a Dorminant Retailer," *Marketing Science*, 28 (2), 309-319.

Farris, Paul, James Olver, and Cornelis De Kluyver (1989), "The Relationship Between Distribution and Market Share," *Marketing Science*, 8 (2), 107-128.

Fieuws, Steffen and Geert Verbeke (2006), "Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles," *Biometrics*, 62 (2), 424-431.

________, ________, Filip Boen, and Christophe Delecluse (2006), "High Dimensional Multivariate Mixed Models for Binary Questionnaire Data," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55 (4), 449-460.

Forman, Chris, Anindya Ghose, and Avi Goldfarb (2009), "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 55 (1), 47-57.

________, ________, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19 (3), 291-313.

________, Avi Goldfarb, and Shane Greenstein (2005), "How did Location Affect Adoption of the Commercial Internet? Global Village vs. Urban Leadership," *Journal of Urban Economics*, 58 (3), 389-420.

Ghose, Anindya, Michael D. Smith, and Rahul Telang (2006), "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research*, 17 (1), 3-19.

Glaeser, Edward L., Jed Kolko, and Albert Saiz (2001), "Consumer City," *Journal of Economic Geography*, 1 (1), 27-50.

Goolsbee, Austan (2000), "In a World Without Borders: The Impact of Taxes on Internet Commerce," *Quarterly Journal of Economics*, 115 (2), 561-576

________ and Peter J. Klenow (2002), "Evidence of Learning and Network Externalities in the Diffusion of Home Computers," *Journal of Law and Economics*, 45 (2), 317-342.

Greene, William (2008), *Econometric Analysis*, Prentice Hall: Upper Saddle River, NJ.

Gueorguieva, Ralitza (2001), "A Multivariate Generalized Linear Mixed Model for Joint Modeling of Clustered Outcomes in the Exponential Family," *Statistical Modeling*, 1, 177-193.

Keeney, Ralph L. (1999), "The Value of Internet Commerce to the Customer," *Management Science*, 45 (4), 533-542.

Knorr-Held, Leonhard and Julian Besag (1998), "Modeling Risk from Disease in Time and Space," *Statistics in Medicine*, 17 (18), 2045-2060.

Kumar, V. and Robert P. Leone (1988), "Measuring the Effect of Retail Store Promotions on Brand and Store Substitution," *Journal of Marketing Research*, 25 (2), 178-185.

Lal, Rajiv and Miklos Sarvary (1999), "When and How Is the Internet Likely to Decrease Price Competition?" *Marketing Science*, 18 (4), 485-503.

Lynch, John G. and Dan Ariely (2000), “Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution,” *Marketing Science*, 19 (1), 83-103.

Michener, Ron and Carla Tighe (1992), “A Poisson Regression Model of Highway Fatalities,” *American Economic Review*, 82 (2), 452–456.

Overby, Eric and Sandy Jap (2009), “Electronic and Physical Market Channels: A Multi-Year Investigation in a Market for Products of Uncertain Quality,” *Management Science*, 55 (6), 940-957.

Reibstein, David J. and Paul W. Farris (1995), “Market Share and Distribution: A Generalization, a Speculation, and Some Implications,” *Marketing Science*, 14 (3), 190-202.

Sinai, Todd and Joel Waldfogel (2004), “Geography and the Internet: Is the Internet a Substitute or a Complement for Cities?” *Journal of Urban Economics*, 56 (1), 1-24.

Tang, Christopher S., David R. Bell, and Teck-Hua Ho (2001), “Store Choice and Shopping Behavior: How Price Format Works,” *California Management Review*, 43 (2), 56-74.

Thum, Yeow Meng (1997), “Hierarchical Linear Models for Multivariate Outcomes,” *Journal of Educational and Behavioral Statistics*, 22 (1), 77-108.

Waldfogel, Joel (2003), “Preference Externalities,” *RAND Journal of Economics*, 34 (1), 557-568.

________ (2007), *The Tyranny of the Market: Why You Can't Always Get What You Want*, Harvard University Press: Cambridge, MA.

Table 1: Preference Minorities, Shelf Space, and Assortment

Retailer Type	**Proportion of HHs with Babies**	**Shelf Space** (width)[1]	**Assortment** Number of SKUs[2]				
			Total	Pampers	Huggies	Luvs	Seventh Generation
Fresh Grocer Supermarket							
Store 1	.106	10 ft	28	9	15	4	0
Store 2	.155	20 ft	50	21	19	8	2
Store 3	.163	28 ft	63	28	25	10	0
Walmart							
Store 1	.139	35 ft	58	24	25	9	0
Store 2	.199	50 ft	82	43	28	11	0

Notes

[1] The retail chains we visited use the same-sized shelves across multiple locations. Within a chain, shelf height and depth is identical; hence we provide only the width information.

[2] Each brand has potentially several variants, e.g., Pampers produces Pampers Baby Dry, Pampers Swaddlers, Pampers Swaddlers Sensitive, and Pampers Cruisers. Moreover, sizes range from “preemie” and newborns to size 6 or 7, and the number actual diapers per package can also vary.

Table 2: Summary Statistics

(a) Dependent Variables: Category Buyers, Category Orders, and Brand Sales

H_1 and H_2:	**Mean**	**SD**	**Median**
Number of Buyers	13.097	16.249	8
Number of Repeat Orders	26.940	58.256	11

H_3 and H_4:	**Mean**	**SD**	**Median**	**Correlation**		
				Pampers	Huggies	Luvs
Number of Pampers Packages	104.409	216.504	38	---		
Number of Huggies Packages	28.653	63.337	9	.815	---	
Number of Luvs Packages	11.657	22.907	0	.307	.287	---
Number of Seventh Generation Packages	22.936	64.446	0	.558	.561	.158

(b) Independent Variables

Variables	**Mean**	**SD**
Preference Isolation		
Number of Total Households	5620.390	3435.880
Number of Households with Babies Aged Less Than 6 Years Old	868.978	541.519
PM Index = [1 - Percentage of Households with Babies]	.837	.054
Online Shopping Costs		
Price: Offline Sales Tax Rate (%)	5.497	3.001
Waiting Time: 1-Day Shipping (1=Yes, 0 = No)	.190	.392
Waiting Time: 2-Day Shipping (1=Yes, 0 = No)	.199	.399
Waiting Time: 3-Day Shipping (1=Yes, 0 = No)	.327	.469
Waiting Time: 2^{nd} Warehouse Led to 1-Day Shipping (1=Yes, 0 = No)	.030	.170
Waiting Time: 2^{nd} Warehouse Led to 2-Day Shipping (1=Yes, 0 = No)	.104	.305
Convenience (H_1 and H_2): Distance to Nearest Supermarket	1.754	2.048
Convenience (H_3 and H_4): Distance to Nearest Supermarket Selling SG	5.145	6.730
Convenience (H_3 and H_4): Distance to Nearest Supermarket with No SG	1.891	2.131
Convenience: Distance to Nearest Discount Store	5.880	5.242
Convenience: Distance to Nearest Warehouse Club	11.476	12.017
Market Potential		
Local Presence of Stores Selling Baby Accessories	.486	1.289
Geo-Demographic Controls		
Percentage of Population Aged 20 to 39 Years Old	.280	.063
Percentage with Bachelors and/or Graduate Degree	.529	.167
Percentage of Female Population in Labor Force	.556	.083
Percentage of Households Below the Poverty Line	.100	.078
Percentage of Blacks	.103	.176
Percentage of Apartment Buildings with 50 Units or More	.034	.068
Percentage of Homes Valued at $250,000 or More	.148	.210
Annual Population Growth Rate from 2000 to 2004	.014	.020
Population Density (thousands in square miles)	1.880	3.719

Table 3: Category-Level Demand Estimates

	Buyers		Repeat Orders	
	Estimate	SE	Estimate	SE
α_0,Model Intercept	-10.256*	.279	-11.563*	.536
H_1: Preference Isolation				
β_1, $PM_{z(m)}$ = [1 - Percentage of Households with Babies]	4.479*	.306	5.644*	.593
H_2: Preference Isolation and Price Sensitivity				
β_3, $PM_{z(m)} \times TAX_{z(m)}$	-.121*	.045	-.162+	.088
Online Shopping Costs				
Price: $TAX_{z(m)}$ = Offline Sales Tax Rate (%)	.113*	.038	.164*	.074
Waiting Time: 1-Day Shipping	.757*	.060	1.230*	.102
Waiting Time: 2-Day Shipping	.363*	.046	.568*	.078
Waiting Time: 3-Day Shipping	.213*	.041	.348*	.070
Waiting Time: 2nd Warehouse Led to 1-Day Shipping	.191*	.075	.401*	.128
Waiting Time: 2nd Warehouse Led to 2-Day Shipping	.128*	.053	.350*	.091
Convenience: Distance to Nearest Supermarket	-.008*	.004	-.024*	.007
Convenience: Distance to Nearest Discount Store	.017*	.002	.021*	.003
Convenience: Distance to Nearest Warehouse Club	.005*	.001	.007*	.001
Market Potential				
Local Presence of Stores Selling Baby Accessories	.007+	.004	.016*	.007
Geo-Demographic Controls				
Percentage of Population Aged 20 to 39 Years Old	2.443*	.136	2.776*	.264
Percentage with Bachelors and/or Graduate Degree	1.677*	.068	1.823*	.128
Percentage of Female Population in Labor Force	-.149	.128	.251	.235
Percentage of Households Below the Poverty Line	-2.551*	.179	-3.416*	.318
Percentage of Blacks	-.274*	.054	-.049	.101
Percentage of Apartment Buildings	.722*	.104	.720*	.207
Percentage of Homes Valued at $250,000 or More	.856*	.047	1.640*	.094
Annual Population Growth Rate from 2000 to 2004	10.114*	.345	10.281*	.676
Population Density (thousands in square miles)	.018*	.002	.025*	.004
Variances				
θ	6.290*	.161	1.096*	.020
τ^2	.201*	.013	.326*	.023
-2LL	52,004		65,645	

Notes
* indicates significance at $p < .05$ and $^+$ indicates significance at $p < .10$.

Table 4: Brand-Level Demand Estimates

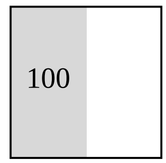
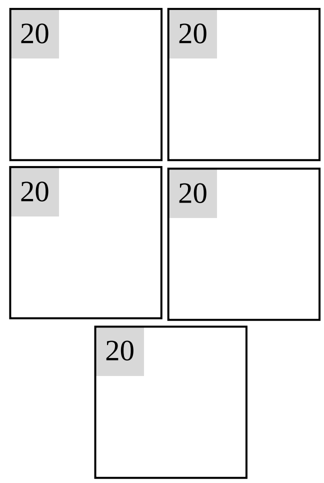
	Popular Brands			Niche Brand
	Pampers	**Huggies**	**Luvs**	**Seventh Generation**
	Estimate	Estimate	Estimate	Estimate
Intercept	-9.180*	-9.854*	-9.655*	-16.094*
H_3: Preference Isolation				
β_1, $PM_{z(m)}$ = [1 - %age of Households with Babies][1]	4.521*	4.164*	3.328*	7.741*
Online Shopping Costs				
Price: $TAX_{z(m)}$ = Offline Sales Tax Rate (%)	.164*	.228*	.179	-.108
Price: $PM_{z(m)} \times TAX_{z(m)}$	-.198*	-.244*	-.201	.128
Waiting Time: 1-Day Shipping	1.035*	1.103*	.838*	1.720*
Waiting Time: 2-Day Shipping	.558*	.517*	.539*	.545*
Waiting Time: 3-Day Shipping	.280*	.387*	.483*	.404^{+}
Waiting Time: 2nd W'house Led to 1-Day Shipping	.403*	.394*	.522*	1.082*
Waiting Time: 2nd W'house Led to 2-Day Shipping	.112	.416*	.195	1.081*
Convenience: Distance to Nearest SG Store	.001	-.002	-.001	.001
Convenience: Distance to Nearest Supermarket	-.012	-.024^{+}	-.007	-.010
Convenience: Distance to Nearest Discount Store	.005	.020*	.004	.041*
Convenience: Distance to Nearest Warehouse Club	.008*	.008*	.008*	-.002
Market Potential				
Local Presence of Stores Selling Baby Accessories	.013	.004	-.004	.013
Geo-Demographic Controls				
Percentage of Population Aged 20 to 39 Years Old	1.371*	.412	.392	3.915*
Percentage with Bachelors and/or Graduate Degree	2.108*	1.950*	1.064*	3.981*
Percentage of Female Population in Labor Force	.788*	.728*	2.186*	1.369^{+}
Percentage of Households Below the Poverty Line	-1.781*	-2.267*	-.767	-1.700
Percentage of Blacks	-.721*	-.424^{+}	-.901*	.457
Percentage of Apartment Buildings	2.193*	2.032*	2.245*	1.179^{+}
Percentage of Homes Valued at $250,000 or More	1.667*	1.324*	-.210	.985*
Annual Population Growth Rate from 2000 to 2004	11.503*	9.351*	6.368*	11.578*
Population Density (thousands in square miles)	.014*	.022*	.012^{+}	.017^{+}
Variance				
θ	.724*	.427*	.210*	0.186*
τ	.344*	.309*	.300*	0.671*
r_{12} (Pampers, Huggies)	.815*			
r_{13} (Pampers, Luvs)	.307			
r_{14} (Pampers, Seventh Generation)	.558*			
r_{23} (Huggies, Luvs)	.288			
r_{24} (Huggies, Seventh Generation)	.561*			
r_{34} (Luvs, Seventh Generation)	.561*			

Notes

* indicates significance at $p < .05$ and $^{+}$ indicates significance at $p < .10$.

[1] The estimate of the PM index for the niche brand, Seventh Generation, is significantly larger ($p < .01$) than those for the national brands, Pampers, Huggies, and Luvs, while these three estimates for the national brands are not significantly different from each other.

Figure 1: Preference Isolation and Geographic Differences in Shelf Space Allocation

Target Population	Total Population	Target Population as Proportion of Total[1]	Stores (200 sq ft each) and Category Shelf Space per Store[2]	Total Shelf Space per market[3]
Market A				
100	200	100 / 200 = 50%		100 sq ft
Market B				
100	1,000	100 / 1,000 = 10%		100 sq ft

Notes

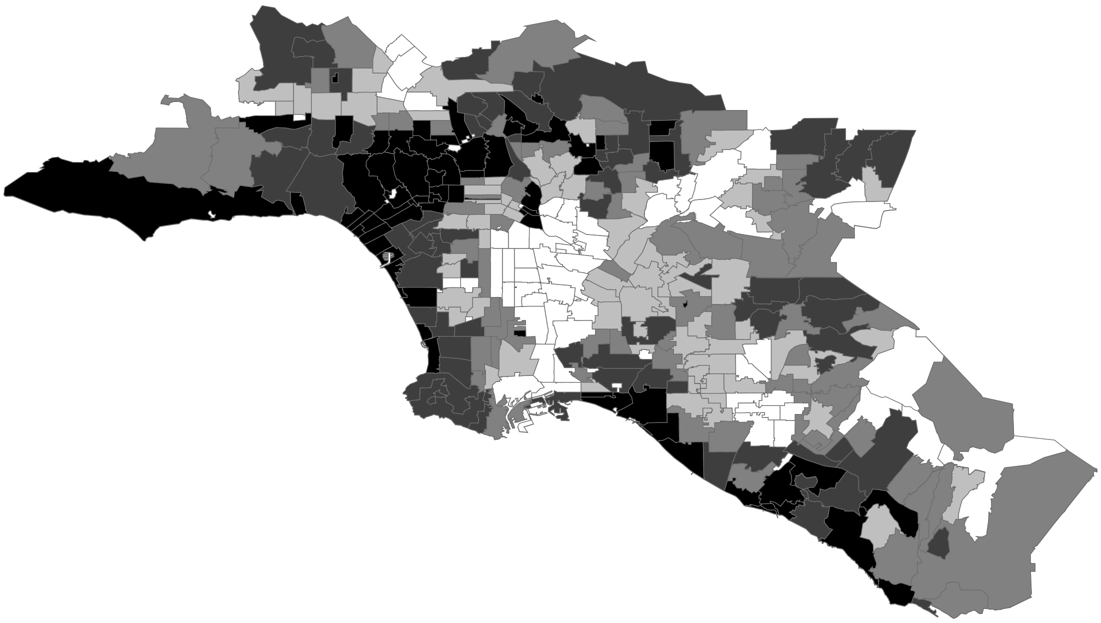
[1] Markets A and B both have 100 residents in the target population, e.g., households with babies. Since Market B has a larger population, the target customers in Market B are, relatively speaking, preference minorities.

[2] Stores are the same size (200 square feet) in both markets, however market B has five times as many stores because it has a larger total population (in the text we use US market data to show that while the number of stores increases with the population size, the size of the stores from a given chain, does not). Stores allocate shelf space to categories in proportion to the size of the target market for that category, i.e., the store in Market A allocates 50% while each store in Market B allocates 10%.

[3] The *aggregate* shelf space allocated to the category is the same in the two markets; however, the *assortment per store* will be much greater in Market A (see Table 1 and Farris, Olver and DeKluyver 1989).

Figure 2: Preference Minorities and the Internet

(a) Los Angeles County: Zip-level Preference Minority (PM) Index



(b) Los Angeles County: Zip Level Orders per Target Household

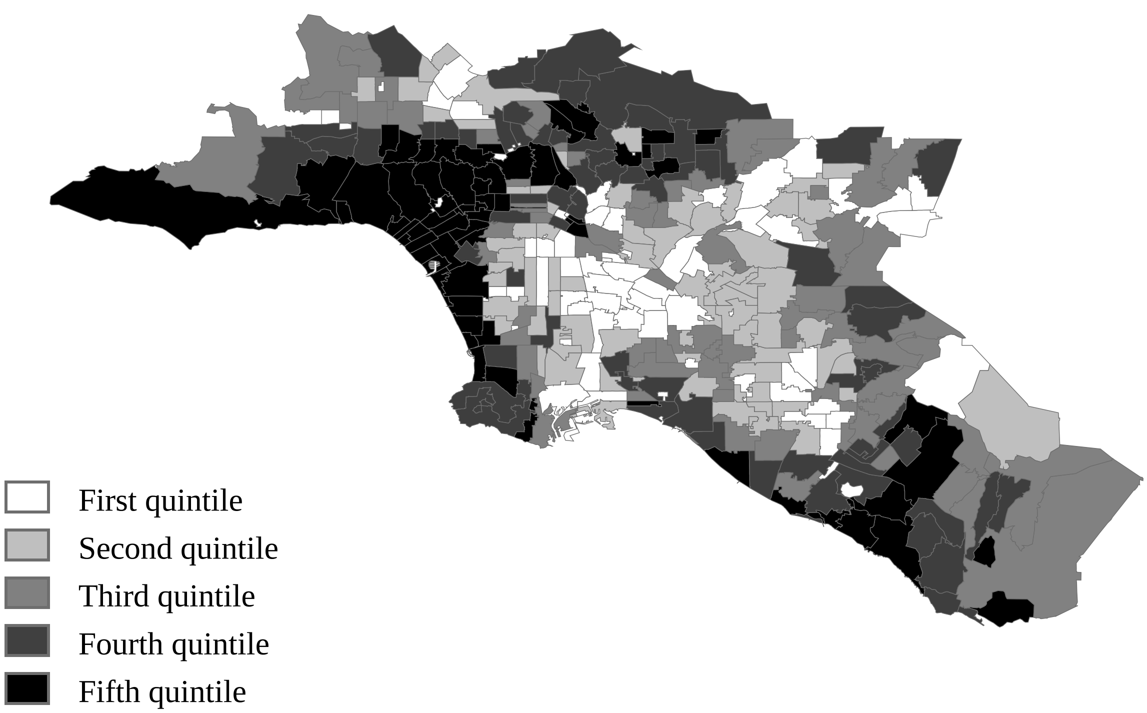
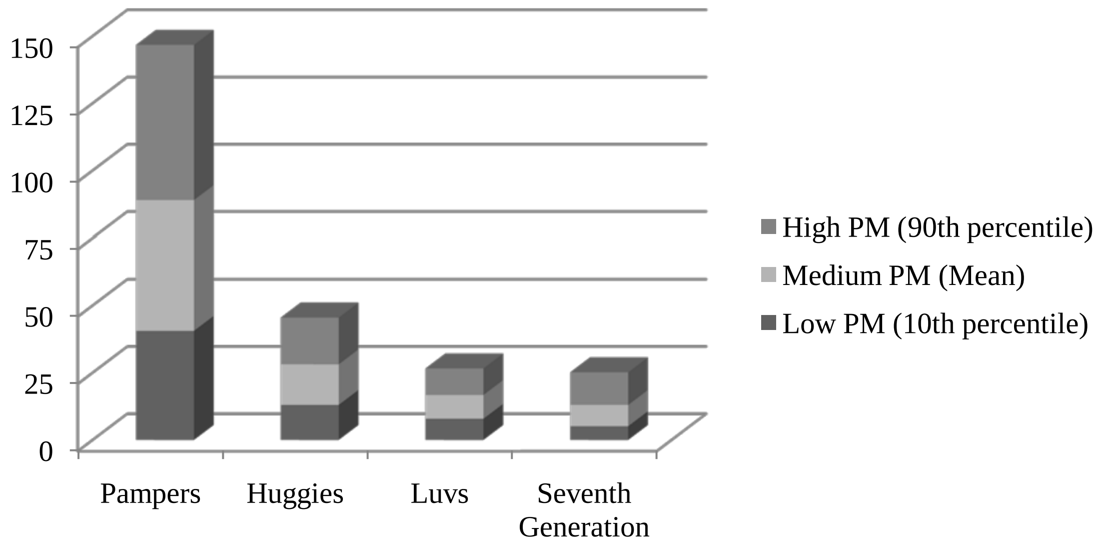


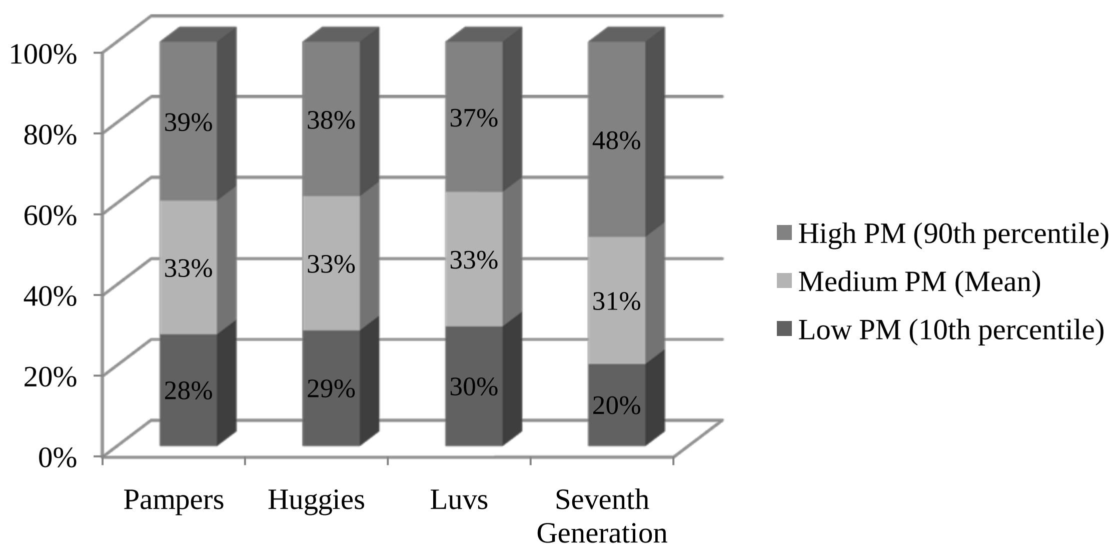
Figure 3: The Contribution of Local Markets to Brand Sales Online

(a) The Long Tail Sales Distribution



Note: We fix the PM Index at the 10th percentile, the mean, and at the 90th percentile. Holding everything else equal, we compute the expected sales for each brand in each of the three prototypical markets.

(b) Decomposition of Brand Sales by Market Type



Note: Preference isolation leads all brands to have higher online sales in markets with higher PM Indices (H_1). For the niche brand (Seventh Generation), substitution from offline to online sales is even stronger (H_3). Hence, preference isolation implies the greatest *relative* proportion of sales come from high PM markets and the effect is amplified for niche brand (H_4). The niche brand sales decomposition in high versus low PM markets is 48:20, or about 2.5:1.