December 2006

The Impact of Location on Consumer Purchases in Electronic Markets

Anindya Ghose
New York University

Avi Goldfarb
University of Toronto

THE IMPACT OF LOCATION ON CONSUMER PURCHASES IN ELECTRONIC MARKETS

IT for Underserved Communities

Chris Forman
Tepper School of Business
Carnegie Mellon University
cforman@andrew.cmu.edu

Anindya Ghose
Stern School of Business
New York University
aghose@stern.nyu.edu

Avi Goldfarb
Rotman School of Management
University of Toronto
agoldfarb@rotman.utoronto.ca

Abstract

The Internet has been thought of as a technological advancement that will bridge the geographical digital divide and remove the disparities between underserved communities and the rest of the society. In order to examine the existence of underserved communities, we examine how changes in the local supply of goods and services in the offline world changes consumer behavior in technology enabled electronic markets. For example, retail markets for consumer products such as books, music, and videos have traditionally been predominantly local. As local markets increase in size from small towns to large cities, consumers are affected in two ways. First, holding product offerings among retailers fixed, an increase in the size of a location encourages new firm entry that in turn lowers prices and improves service levels. Moreover, larger markets also allow retailers to provide a wider array of product offerings targeted to market segments that would be infeasible in smaller town settings. The emergence of new online retailing channels may act as a substitute for the benefits of urban concentration both by offering lower prices and by providing increased product differentiation for rural consumers. The open question is which of these phenomena are more important: how do consumers use online channels to substitute for offline supply deficiencies, and how does this vary across locations in the United States? We explore this problem using data from Amazon on the top selling books and DVDs for over 8626 unique locations in the US over 10 months between 2005-2006. We show that even controlling for product-specific preferences by location, there are still considerable differences in the responses of locations of different population sizes to price and popularity changes of DVDs and books. These can be attributed to differences in local supply conditions as well as other factors such as demographic characteristics and high speed internet penetration.

Keywords: Digital divide, retail competition, geographical location, electronic markets, offline channels, inner communities.

Introduction

It is well documented now that the Internet retailing revolution has established a new distribution channel that represents a fundamental paradigm shift in consumer buying patterns. The rapid growth of alternative retail channels has transformed not only the competitive structure of several industries, but also the way in which consumers shop for products, especially in those categories where purchases are frequent. Information goods such as books, CDs, DVDs, Videos etc, are examples of such products. Like offline retailing, the key components of online retailing are selection, convenience, and price.

Location matters to the use of Internet retailing. People in urban areas buy a different set of products online than people in rural areas. The benefits of online retailing vary by locations due to differences in offline selection, convenience, and price. In this paper, we explore how differences in local supply conditions influence consumer purchases in electronic markets. We show that consumers benefit from electronic markets due to improved price, selection, and convenience, and further show how each of these benefits varies across locations.

Since the inception of online retailing in the late 1990s, selection on the Web has increasingly become broader and deeper. Internet retailers have nearly unlimited "virtual inventory" through centralized warehouses and drop shipping agreements with distributors (Mendelson and Meza 2002). This enables them to offer convenient access to a larger selection of products than can brick-and-mortar retailers. For example, small stores stock approximately 20,000 unique titles, large independent booksellers stock approximately 40,000 unique titles, Barnes & Noble and Borders superstores stock approximately 100,000 unique titles (Brynjolfsson et al. 2003). Large differences in product variety are also seen in music, movies, and consumer electronics products. Even Wal-Mart Supercenters, which can be up to 230,000 square feet in size, only carry one-sixth of the number of SKUs that are carried by Walmart.com (Owen 2002). Amazon has 40,000 unique DVD titles, while a typical Blockbuster has about 5000 titles and Netflix has about 60,000 DVD titles.

Shopping convenience plays an important role in internet commerce (Tokzadeh and Dhillon 2002). Unlike traditional retail and most mail order companies, the Internet is available for transactions at all times. In addition, features like one-click ordering and overnight delivery translate into significant time and energy savings for consumers. For example, customers for online grocery retailers typically tend to be time-starved shoppers who value convenience and hail from mid/upscale suburbs. Brynjolffson et al. (2003) note that an average order placed on barnesandnoble.com took three minutes to place, and arrived in three days. In contrast, the Barnes & Noble order took nearly one hour to place, took eight days to arrive, and was approximately $6 more expensive than the online order. Given that an average consumer lives about 5.4 miles from the nearest bookstore, online channels offer a significant amount of convenience to consumers. Thus, knowledge of consumer purchase patterns according to geographical locations could be very useful in firms' marketing mix strategies. For example, online retailers could target their promotions in specific geographical areas where customers are more likely to derive value from their services as noted by Jank and Kannan (2006).

Prior work has also shown that online prices are often significantly lower than offline prices (e.g. Brynjolfsson and Smith 2000). The difference between online and offline prices that consumers face varies significantly across locations due to differences in local supply: locations with more stores are expected to have lower prices, other things equal.

Despite a wealth of research on electronic commerce to the best of our knowledge, no prior research has measured how these competing benefits shape online consumer buying behavior across geographic locations. We explore how these benefits lead to systematic differences in product purchases across geographical regions. Figures 1a and 1b provide a summary of our results. It shows that people in small communities buy more popular products online than people in large cities. Furthermore, people in small communities buy newer products and slightly more expensive products. However, these differences in online purchases across locations may be due to differences in local supply conditions or differences in consumer preferences across locations. In this paper, we provide a framework for identifying between these alternative explanations.

Utilizing a panel data set of online purchases of books, and DVDs purchased by consumers across urban and rural locations in the US, we propose an econometric study to examine how location shapes consumer use of online channels. Our empirical strategy addresses an important problem that has been inadequately addressed in statistical

e-commerce research: how consumers substitute between online and offline channels. We develop a methodology for understanding online/offline substitution by combining online purchase data with offline data on demographics and local channels. In particular, controlling for consumer preferences, we examine whether rurally based consumers purchase systematically more popular or less popular products than urban customers. This in turn demonstrates whether rural consumers use online channels to obtain better product offerings or to obtain lower prices on mass market items, relative to urban consumers. Since the broader aim of this study is to explore if there is a digital divide between different groups of society, we use socio-economic factors such as age, income and education as independent variables. By comparing the technology adoption among what is known as the "risk groups" to the adoption among the population average and analyzing consumer buying behavior, we aim to get some insights into the digital divide that exists in the US today. Our ongoing work aims to analyze these implications in greater detail by identifying specific regions or regional clusters which are affected by the lag in the penetration of advanced Internet access such as broadband and other communication technologies.

Our paper also had some implications for retailers who would like to better understand their consumers in order to elicit higher response rates from them. Indeed, using spatial data is particularly appealing in multi-channel situations when traditional channels, and direct marketing channels are used along with the online channel to target customers (Jank and Kannan 2006).

The remainder of the paper is organized as follows. Section 2 provides background of the prior literature. Section 3 describes the data. Our empirical strategy is explained in Section 4, and results are presented and discussed in Section 5. Section 6 concludes.

Related Literature

This paper is related to several streams of research. First, it contributes to recent literature on the Digital Divide. The literature on the Digital Divide is too extensive to survey here, for a helpful summary see Dewan and Riggins (2005). Recent studies exploring this phenomenon on an international level include among others, Kauffman and Techatassanasoontorn (2005) and Lam and Lee (2006**).** Recent work in the Digital Divide literature has sought to quantify the benefits of Internet access for economic growth (Gillet et al 2006), and quantify the implications for underserved communities. In this paper, we explore one particular benefit of Internet access: the ability to substitute for offline retailers and provide better price, convenience, and selection for consumers. We also show how these benefits vary according to broadband access, presenting evidence on one particular economic implication of the digital divide.

Our paper is also related to recent research that examines the potential for Internet technology to reduce the costs associated with distance. Forman et al. (2006) show that new communications technology may be most valuable for rural businesses. Sinai and Waldfogel (2004) show that minority individuals and individuals who are geographically isolated are more likely to be connected to the Internet. Moreover, they show that geographically isolated individuals spend more money online. While this work demonstrates that information technology in general and the Internet in particular can lower the costs associated with distance, it does not address the primary question in our paper: how IT creates benefits to consumers in terms of improved price, selection, and convenience.

A number of recent papers have examined how various parameters in retail settings vary across geographical markets by incorporating spatial structure into models of aggregate demand/sales (Bronnenberg and Mahajan 2001, Jank and Kannan 2006). These papers suggest that spatial data captures not only the geographical variations in supply side factors (stores related variations in locations, products, layout, service, order-of-entry, etc.) but also variations in demand side factors (geographical characteristics and customer characteristics impacting preferences and consumption patterns), each of which can influence consumer buying behavior. Our paper provides a unique methodology of identifying how supply factors influence consumer buying behavior separately from demand side factors such as consumer characteristics.

Empirical papers that explore the determinants of local inequalities show that, even at the regional level, they depend on (differences in) economic, social and demographic indicators. In particular, differences in the spatial diffusion of ICTs have been attributed to differences in technological levels, infrastructural endowments and local

spillover effects (Audretsch and Feldman, 1996). Recent work has pointed out that differences in spatial diffusion may be influenced by pure spatial factors, rather than by heterogeneity in endowments at regional or local level. For instance, Nunes (2004), investigating the geography of top level domain names in Europe, has proposed that in some cases Internet contributes to reinforce the tendency to territorial disintegration, promoting geographic disparities in a more pronounced way than is the case in the real economy space.

Second, this paper contributes to recent research that demonstrates the welfare benefits that online channels provide to consumers by lowering the search costs of obtaining hard-to-find books (Brynjolfsson et al. 2003) and by establishing active resale markets (Ghose et al. 2005). Using data from Amazon (Ghose et al. 2006) show that IT-enabled online markets, have a strong positive first-order impact on total welfare and consumers in particular. We also contribute to existing research that examines consumer sensitivity to online and offline prices. While there is a burgeoning literature studying how e-commerce has affected prices and price dispersion (Brown and Goolsbee (2002); Brynjolfsson and Smith (2000)), and how one can use ecommerce data for demand estimation to highlight consumer sensitivity to prices (Ghose and Sundararajan 2006), much less work has looked at how the diffusion of the Internet has influenced the distribution of product sales in online channels and how that distribution has been affected by the entry and exit of retail stores in offline markets.

Theory

There are a number of possible reasons why the types of products purchased online may vary across locations. First, consumers in different locations may simply have different underlying preferences. For example, rural consumers may prefer more popular items independent of local supply characteristics. If the differences are simply based on preferences, we would expect that controlling for location-specific product preferences would eliminate any differences between urban and rural customers.

The other possible reasons why urban and rural customers may have different online behavior reflect substitution between purchasing from local offline retailers and purchasing online. In particular, the characteristics of local offline retailers vary from location to location. Some locations have many specialty retailers who sell a selection of products at full price. Some locations have very large specialty stores that sell thousands of products in the same category. And some locations have large discount stores like Walmart that sell a very small number of products in a given category at relatively low prices. There is considerable variation by location in the relative availability of these three kinds of retailers to buy offline. The offline options available to consumers affect their online choices. We test these predictions using our dataset.

We consider a market consisting of two firms producing a differentiated product in a two-stage non-cooperative game. In Stage 1, firms i and j simultaneously choose optimal prices (p_i, p_j) simultaneously. Then consumer demand is realized in Stage 2. We consider a subgame perfect equilibrium of this game using backward induction. One firm is a brick and mortar (offline) retailer (labeled as *i*) and the other firm is an internet only (online) retailer (labeled as *j*). We assume that the demand facing each product is linear in self and cross-price effects. The demand function for the offline firm is given by

$$q_i = 1\text{-}A\text{-}\ b_1p_i + b_2p_j \qquad j=1,2\ ;\ i \neq j \qquad (1)$$

and the demand function for the online firm is given by

$$q_j = A\text{-}\ b_1p_j + b_2p_i \qquad j=1,2\ ;\ i \neq j \qquad (2)$$

where $0 \leq b_2 < b_1$. Here b_1 and b_2 are the own price and cross-price effects.

From this point onwards, we will continue to designate the firm under consideration as firm *i* and its competitor as firm *j*. As well, in describing the behavior of both firms in terms of a system of equations, as expressed in (1), we will drop the qualification, $i,j = 1,2;\ i \neq j$ for brevity. The parameter *A* represents the intercept of the demand curve which can be construed as the distance to the offline store from a consumer's location (derived from a Hotelling like model). Then, as *A* increases, consumers face increased transportation costs of traveling to the offline store, resulting in an increased likelihood of purchasing from the online channel.

Each firm sells two products, labeled 1 and 2, which differ based on the popularity of the product. Without loss of generality, let 1 be the unpopular (rare) product and 2 be the popular product. Let *a* be the parameter representing

the popularity of the product. Thus as *a* increases, product 1 becomes relatively less popular and product 2 becomes relatively more popular. Let *B* denote the probability of availability (expected availability) for the unpopular product in the offline store.

Let the demand of the offline firm for product 1 be denoted by D_1 and the demand of the offline firm for product 2 be denoted by D_2. To reflect the fact that offline retailers have capacity constraints from limited shelf-space or floor space, we incorporate the possibility that they face a trade-off from stocking popular versus unpopular products. Let this parameter be *k*. Hence, we have the following constraints for unpopular and popular products, respectively.

$$Bk = D_1 \quad (3)$$
$$(1-B)k = D_2 \quad (4)$$

Let the Lagrangian constraints for the limited capacity of the offline retailer for stocking unpopular and popular products be given by λ_1 and λ_2. We assume that the online retailer has no such constraints due to its virtually unlimited capacity to stock products.

Therefore, the profit function of the offline firm is given by

$$R \underbrace{(1-A- b_1p_{i1} + b_2p_{j1})}_{D_1} p_{i1} + (1-R) \underbrace{(1-A- b_1p_{i2} + b_2p_{j2})}_{D_2} p_{i2} +$$
$$\lambda_1 (Bk -D_1) + \lambda_2 ((1-B)k -D_2)$$

which upon relevant substitution becomes equal to the following expression

$$\pi_1 = R \underbrace{(1-A- b_1p_{i1} + b_2p_{j1})}_{D_1} p_{i1} + (1-R) \underbrace{(1-A- b_1p_{i2} + b_2p_{j2})}_{D_2} p_{i2} + \lambda_1 (b_1p_{i1} - b_2p_{j1} + Bk -1+A) + \lambda_2 (b_1p_{i2} - b_2p_{j2} +(1-B)k -1+A) \quad (5)$$

Similarly the profit function of the online firm is given by

$$\pi_2 = a \underbrace{(A- b_1p_{j1} + b_2p_{i1})}_{D_3} p_{i1} + (1-a) \underbrace{(A- b_1p_{j2} + b_2p_{i2})}_{D_4} p_{i2} \quad (6)$$

Here D_1 and D_3 represent the expected demand curves for each firm for product 1, while D_2 and D_4 represent the expected demand curves for each firm for product 2. The above equations are then plugged into the profit functions. This gives us six reaction functions which are derived by setting the six first order conditions to zero. The solutions to these reactions functions will give us the optimal parameter values for $\lambda^*_1, \lambda^*_2, p^*_{i1}, p^*_{j1}, p^*_{i2},$ and p^*_{j2}
Hence, in equilibrium the realized demand curve expressions are given by the following equations

$$M_1 = akB \quad (7)$$

$$M_2 = (1-a)k(1-B) \quad (8)$$

$$M_3 = a\left(\frac{b_1(A(b_1-b_2)+b_2(1-Bk))}{2b^2_1 - b2^2}\right) \quad (9)$$

$$M_4 = (1-a)\left(\frac{b_1(A(b_1-b_2)+b_2(1-(1-B)k))}{2b^2_1 - b2^2}\right) \quad (10)$$

Based on the above equations and the profit functions, we are able to derive the following three hypotheses:

Proposition 1: Online purchasing for convenience: *As distance to offline stores, A, decreases, the proportion of popular products being bought online also decreases.*

Proposition 2: Online purchasing for selection: *As availability in offline markets, B, decreases, the proportion of rare products (less popular products) bought online increases.*

Proposition 3: Online purchasing for prices: *As distance to offline stores, A, decreases due to the entry of a store, the relative price changes and hence the proportion of expensive products bought online increases.*[1]

A detailed intuition for these propositions is displayed in figures 4a, 4b and 4c.

Data

The data that we use for this study come from the Web pages on "Purchase Circles" on the Amazon.com Web site. We used a JAVA spider to extract and parse data from Amazon's Web site. Purchase Circles are specialized best-seller lists. The pages denote the top-selling books, music, and DVDs across large and small towns throughout the US.[2] For many towns it also displays the top selling electronics products, video and toys. Books, DVDs, and music comprise an important share of Amazon's "media" business which accounts for 75% of all sales. The total revenues of Amazon from the media business grew 18% in 2005 and was approximately equal to $3.05 billion.[3]

We collect weekly and fortnightly data on purchase circles over a period of ten months from April 2005 to January 2006 which is then aggregated on a monthly basis. As Amazon.com is by far the largest online retailer (for instance, it has 70% of the market share in the book market-Forrester Research 2002), this data set provides an excellent way of measuring the use of online channels across locations within the US. The Purchase Circles are organized in multiple layers- first, by state and then within a state, by town. For larger towns, they are also organized by suburb or county. Thus, the number of geographical locations that are listed for any given state is very large: for example, the state of Pennsylvania alone lists 271 towns. As a result, we have a large panel data set. Though previous studies have used data from the Amazon in the past, we are the first to use the data available through Purchase Circles.

For each town, Amazon provides a list of top 10 sellers in the town for each product category.[4] For larger towns, it provides the top 20 sellers for each product category. Amazon also provides on its Web site the national sales rank within that product category (books, music, etc.) for each product that it sells. So, for example, we know that in April 2006 the product Angels and Demons by Dan Brown was ranked #27 in books nationally while it is ranked #5 in Las Vegas, Nevada, #6 in Great Falls, Montana and #3 in Brooklyn, New York. Similarly, while Soul Sessions by Joss Stone is ranked # 92 nationally in music, it is ranked # 9 in Anchorage, Alaska, # 2 in Franklin, Michigan and #15 in Boulder, Colorado. We match the list of top ten sellers for each town to Amazon's national sales rank to determine the sales rank for top sellers in each town. This local sales rank will be our primary dependent variable. Descriptive statistics for DVDs are provided in Table 1 and those for Books are provided in Table 2.

Product Characteristics: For every product sold on its Web site, Amazon provides extensive details about it. For instance, for any given book which is listed on the Purchase circle, we can go to the next inner layer and collect data on the specific characteristics of the book, based on its ISBN number. These include the product (Book and DVD) list and retail prices, release date of the product in the market, the average customer rating, the number of reviews, the number of used copies available for sale, the product type and genre, and so on. This enables us to determine what product characteristics, if any, are uniquely attributable to location, both in rural and urban areas.[5] In this study, we focus on the DVD data. We also present a small number of results for books to ensure robustness.

The "Relative Price" variable is the difference between the offline price and the Amazon retail price (normalized by the offline price). We use information on the product's list price as a proxy for the offline price. Amazon also provides information on the offline price that the product is being sold at, for instance through its affiliate Borders, and our analysis reveals that there is very high one-to-one mapping between the offline price and the list prices in categories such as books and DVDs. The "Very Popular" variable was defined as a product which was in the top 25

[1] Proofs of all Propositions are omitted for brevity but are available from the authors upon request.
[2] Henceforth, we use the word towns to refer to small and large cities, as well as small towns.
[3] http://fonerbooks.com/booksale.htm
[4] For some towns, information on how popular the product is relative to the US is also provided. Though the discussion below is based on the list of best sellers, we plan to use both sources of information.
[5] In addition to highest selling products by location, Amazon also publishes a list of top 10 or top 20 products, which is "uniquely popular" in a given town, as compared with the rest of the country, at any given period of time. We do not currently use this information, but plan to do so in future research.

books overall at Amazon measured by the sales rank. The main dependent variable is a dummy which is 1 if the product is present in the top 20 for a given location and time.

Local Demand Characteristics (Demographics): To obtain data on location size and demographic characteristics, we match our data set to US Census data. Data on population size by town, county, Metropolitan Statistical Area (MSA), and state are provided annually by the Census Bureau, while demographic data are available at various levels of aggregation in the decennial Census.[6] In order to code the broadband penetration across different regions, we used a dataset from FCC that listed the number of broadband firms in different zipcodes.[7]

Local Supply Characteristics (Stores): We have data on the change in the total number of stores for a given location from County Business Patterns (CBP). CBP data contain detailed data at the industry level such as the total number of establishments (stores) as well as their size distribution by number of employees. Stores could be classified as discounters, big specialty stores (e.g. Barnes and Noble), and small specialty stores (e.g. local booksellers). The size of these stores matters when we consider the interplay between selection and price in offline markets. Changes in the size distribution provide critical information about heterogeneous effects of online retailing in a geographical location. Table 3 summarizes the trade-off that consumers face when they shop in offline channels. We have also collected data on entry and exit of stores of Wal-Mart, Target, Barnes and Noble, Borders and Best Buy corporations. These data were collected either through information on press releases or from the companies directly. We use this data to control more fully for location-specific preferences. Figures 3a and 3b show how the presence of discounters and big specialty stores has an affect on the main variables of interest.

Econometric Framework

We seek to understand the differences in online purchase behavior across locations. In particular, we are interested in how local supply of retailers influences the decision to buy online. We use the information on the top selling products in each location to determine how the characteristics of the location influence the online/offline purchase decision. In particular, we view the decision to buy a particular item online as the consequence of two separate decisions: (1) the decision to buy the item and (2) the decision to buy online rather than offline. Conditional on buying an item, the consumer's online/offline purchase decision will depend on the local supply of the item. Our analysis controls for location-product fixed effects. The identification comes off how changes in product characteristics over time influence the interaction between location and product characteristics. Using this framework, we study how geographic variations in offline selection, convenience, and price influence online product choice.

The rich panel data allow us to estimate our regressions at the location-product-month level. In particular, we estimate the following linear regression:

$$(1) \qquad Top10Rank_{ijt} = Z_{it}\alpha + X_j\beta + Z_{it} \cdot X_j\gamma + \mu_{ij} + \varepsilon_{ijt}$$

where $Top10Rank_{ijt}$ is a dummy variable for whether product i is in the top 10 in location j for month t; Z_{it} are the attributes of product i for month t including price, overall sales rank, whether it is in the top 25 most popular items in the category, average rating, number of reviews, and elapsed time since launch;[8] X_j are the attributes of location j including per capita income, percent college graduate, population, percent under age 24, broadband availability, number of specialty stores, number of discounters, and number of big specialty stores; μ_{ij} is a product-location fixed

[6] The data were generated by using the county where the location had the most zip codes. In cases where there was a tie, we averaged across counties. These are counties where one county has zero and one county has one.

[7] In cases where the zipcode was not listed, we coded the "Broadband" variable as equal to 0. Basically, the broadband variable is defined as the mean number of broadband firms across all zipcodes in the county. This implies that there are no counties in the sample where there does not exist at least one zip code with a broadband internet provider.

[8] The price information is missing for a number of products. In these cases, we include a dummy variable indicating a "missing price". Therefore the missing observations do not affect the price coefficients.

effect (meaning that $X_j\beta$ will difference out), and ε_{ijt} is a product-location-month idiosyncratic error term. The main coefficients of interest are γ, the interaction of local supply characteristics and changes in product characteristics.

Our identification strategy employs changes in product characteristics over time to measure how variance in offline supply of retailers influences which products consumers buy. As noted above, our data on store opening (such as Walmart, Target and Barnes and Nobles) will ensure that location-specific price and overall sales rank sensitivity does not drive results. Instead, results will be driven by variation in local supply characteristics

Results

Analysis of DVDs

In this section, we first show that even controlling for product-specific preferences by location, there are still considerable differences in the responses of locations of different population sizes to price and popularity changes of DVDs. We then try to understand why these different locations display different responses in the context of local offline convenience, selection, and prices. We compare locations with very few local retail options with other locations to see how much of the difference in behavior across location size is explained by different types of local retail supply. Finally, we compare the relative importance of convenience, selection, and price in explaining the differences in online purchasing behavior by location. Our hypotheses are summarized in Table 4.

Table 5 examines DVD sales online. Columns 1 and 2 show that population size has an impact on responsiveness to changes in price and popularity. This is a more rigorous test of figure 1. Columns 3 through 6 explore whether controlling for the number of discount stores and the number of large DVD specialty stores affects the differences between cities and rural areas. Columns 3 and 4 show that controlling for the number of discounters in a location effectively eliminates any difference in price response between cities and rural areas.

Table 6 highlights the three main effects after controlling for demographic characteristics. Interestingly, we find some evidence of the *price effect* we expected: the interaction between price and the number of discounters is negative and significantly different from zero.[9] We do find evidence of a *convenience effec*t because the interaction between Very Popular and the number of discounters is negative (more than 50% of DVDs are sold at discounters (Chiou 2005)) in both tables 5 and 6. Columns (5) and (6) in table 5 and column (1) in table 6 show evidence of a *selection effect*: the interaction between Sales Rank and whether there is a large specialty DVD retailer is negative. Still, neither of these effects explains the differences between cities and rural areas in the popularity of the items they buy online. Both cities and rural areas are affected by selection and convenience effects in similar ways, but these effects do not appear to mitigate the overall differences.

We also conduct analysis with the key socio-economic variables to explore the extent of the digital divide. Our preliminary results reveal that the interaction of Sales Rank and broadband penetration is positive. This result is in accordance with theoretical predictions since one would expect that with a decrease in online search costs, lower ranked products become more popular. Moreover, the interaction of Sales Rank with the dummy for the percentage of college educated residents and percentage of people under the age of 24 is negative. Essentially, as the average education of consumers increase, they are more likely to buy rare/unpopular/obscure products. Moreover, towns with higher average per capita income pay are less sensitive to changes in the online price. All this suggests evidence of some kind of a ***digital divide***.

Figure 2 summarizes the DVD results. It shows that much of the difference between cities and rural areas in their responsiveness to changes in DVD prices can be explained by the presence of discount stores. However, the difference between cities and rural areas in their responsiveness to overall product popularity cannot be explained by the presence of large specialty stores.

[9] Note that the sign of the "very popular" variable is negative for DVDs when we consider only the top-25 best selling products as "very popular". If we consider the top-100 DVDs as the threshold for defining the most popular DVDs, then this is actually positive just as expected.

Analysis of Books

Next we discuss the books results using a more robust econometric framework that incorporates entry or exit of actual stores over time in a given location.[10] We began our research with an investigation of the geographic variance in consumer preferences across locations. Figure 1b shows that there is considerable variation in which books are in the top 10 in the online market across locations. The vast majority of products that are counted in the top ten in at least one location are in the top ten only in a small fraction of the total locations. Figure 1b also shows that people in smaller towns buy more popular products online than people in large cities. Furthermore, people in smaller communities buy newer products and slightly more expensive products. Whether this is due to different preferences or different local supply conditions is an empirical question that we discuss below.

Table 7 presents our main results for books using the entry data. In particular, we define the most popular products as those which are in the top 150 while we define the less popular products as those which are in top 1500-15000 range. These different splines help in calibrating the magnitude of the selection and convenience effects.[11] Moreover we also split the sample by city size (in terms of its population) and run regressions with towns having population under 100000, towns whose population lies between 100K-1 million and another with towns whose population exceeds one million.

As expected, the interaction terms in the first three rows are negative and significant. Based on our theoretical model, these results suggest that convenience, selection, and price matter differently to consumers in locations with different local retailer environments. The variation in the availability of local offline retail options is correlated with variation in the types of books consumers buy online. Thus, table 7 shows evidence in favor of all three propositions: convenience, selection, and price.

Note that we find strong evidence of a *convenience effect* because the direct effect of the interaction between "Very Popular" books and the entry of a discount store like Walmart or Target as well as the entry of a specialty store such as Barnes and Noble or Borders is negative. The other coefficients suggest that as products become progressively less popular, the likelihood of them appearing in the top 10 decreases.

Similarly, note that the estimates also show some evidence of a *selection effect* because the interaction between books that are moderately popular or unpopular and the entry of a large discounter such as Walmart or Target or entry of a specialty book retailer such as Barnes and Noble or Borders is negative.

Next we examine the *price effect* and find strong evidence suggesting that local retailer options are related to the price benefits of going online. Note in particular, the sign on the *relative price* variable is positive while the sign on the interaction of relative price with entry of Walmart and Target or the interaction with entry of Borders and Barnes and Noble are both negative. The positive sign on the *relative price* variable implies that as the offline price decreases, there is a decreased likelihood of the book being bought online (leading to its exclusion from the top 10 most popular list). The negative sign on the interaction implies that with the entry of a discount store such as Walmart people are more likely to buy discounted products offline, and only buy expensive products online. Similarly, the entry of a specialty store such as Barnes and Noble also has a similar effect on consumers' increased propensity to buy expensive products through the online channel.

Discussion and Conclusion

A recent report from the U.S. Department of the Commerce (2004) documents that while many more Americans are getting connected to the Internet, there is a widening gap among households of different educational, income, racial,

[10] Results for DVDs using store entry are qualitatively similar and hence we omit them for brevity.

[11] As a robustness check we perform the same analysis by defining the most popular products as those which are in the top 100 and the less popular as those which are in the 1000-10000 range. Note that the results are very consistent across all specifications.

geographic backgrounds and disabled individuals. The result is that many low-income, rural and small-town communities are being left out of this information revolution and are deprived of the sense of economic opportunity it offers. The gap between citizens from different socio-economic backgrounds with regard to their opportunities and abilities to access and use information and communication technologies is regarded as a potential barrier for participation in the information society.

Utilizing a unique panel data set of online purchases of books, and DVDs by consumers across urban and rural locations in the US, we propose an econometric study to examine how location shapes consumer use of online channels. In particular, controlling for consumer preferences, we examine whether consumers with few local retail options purchase systematically more popular or less popular and more or less expensive products than urban customers. This in turn demonstrates how the selection, convenience, and price benefits of the Internet are different for customers in different types of locations.

We show that even controlling for product-specific preferences by location, there are still considerable differences in the responses of locations of different population sizes to price and popularity changes of DVDs and books. These differences can be attributed to differences in local supply conditions including internet penetration. Since product popularity can affect prices and the volume of transactions, higher prices, lower frequency of purchases and fewer units of purchases can all have a detrimental effect on social welfare, particularly if these effects are worse in inner communities and regions which are lagging behind in socio-economic progress.

In sum, this study shows that Internet access can have significant impacts on consumer welfare in terms of improved price, selection, and convenience. This paper shows how use of electronic commerce varies across consumers in urban and rural areas. It demonstrates whether consumers in rural areas use electronic commerce to obtain products that are difficult to find in rural areas, or whether the use of electronic commerce acts a substitute channel for mass market products that are available but not adequately services by existing channels. It contributes to existing research that examines how IT reduces the costs associated with distance. It has implications for managers of traditional physical retailers who wish to understand how online retailers substitute for offline retailers. It also has implications for manager of online retailers who wish to understand the behavior of the consumers that they serve. While online businesses do collect a significant amount of information on their customers in order to mine their purchasing history, and Web site browsing behavior, the use of such data is limited in the case of product categories where purchases are infrequent. In such situations, data about geographical purchase patterns can be very useful to understand customers' choices better.

References

Audretsch D., and Feldman M. "Knowledge Spillovers And The Geography Of Innovation And Production," *American Economic Review,* 86, 1996, 630-640.

Bronnenberg, B. and Mahajan, V. "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables," *Marketing Science*, (20: 3), Summer, 2001, 284-299.

Brown, J. and Goolsbee, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy,* (110:3), 2003, 481-507, 2002

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

Dewan, S., and Riggins, F. "The Digital Divide: Current and Future Research Directions," *Journal of the Association for Information Systems*, 6, 12, December 2005, 298-337.

Forman, C., A. Goldfarb, and Greenstein, S. "How Did Location Affect Adoption of the Commercial Internet: Global Village vs. Urban Leadership," *Journal of Urban Economics* 58(3), 2005, 389-420.

Ghose, A., M. Smith, and Telang, R. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Implications," *Information Systems Research*, 17(1), 2006, 1-19.

Ghose, A., R. Telang, R. Krishnan. "Effect of Electronic Secondary Markets on the Supply Chain," *Journal of Management Information Systems* (22:2), 2005, 91-120.

Ghose, A. and A. Sundararajan 2006. **"**Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges," *Statistical Science*, 22(1), May 2006.

Gillett, Shane, William Lehr, Carlos A. Osorio, and Marvin A. Sirbu. 2006. Measuring the Economic Impact of Broadband Deployment. Report prepared for the U.S. Department of Commerce, Economic Development Administration.

Goolsbee, A. "In A World Without Borders: The Impact of Taxes On Internet Commerce," *Quarterly Journal of Economics* 115(2): 561-576.

Jank, W. and Kannan, P. "Understanding Geographical Markets of Online Firms Using Spatial Models Of Customer Choice," *Marketing Science*, 24 (4), 2006, p.623--634.

Kauffman, R and Techatassanasoontorn, A. "Is There A Global Digital Divide For Digital Wireless Phone Technologies?," *Journal of the Association of Information Systems,* 2005.

Lam, J., and Lee, M. "Digital Inclusiveness: Longitudinal Study of Internet Adoption by Older Adults," *Journal of Management Information System*s, 22, 4, Spring 2006, 177-206

Lazarus and Lipper, J. "The Search for High-Quality Online Content for Low-Income and Underserved Americans: Evaluating and Producing What's Needed," Case Study, 2003. www.childrenspartnership.org

Mendelson, H., and P. Meza. "Amazon.com: Marching toward profitability," Stanford Graduate School of Business, Case EC-25, July 2002.

Nunes F. "The Geography of .Pt Top Level Domain," *Paper presented at the 44th European Congress of the European Regional Science Association,* August 25-29, 2004, Porto.

Owen, T. "Lord of the things," *Bus. 2.0* (March), 2002, 32.

Robinson, J. "The Role of Information Technology in the Economic Development of Inner City Communities," Working Paper, Columbia University, 2000.

Sinai, T. and Waldfogel, J. "Geography and the Internet: is the Internet a substitute or complement for cities?," *Journal of Urban Economics* 56, 2004, 1-24.

Torkzadeh, G. and Dhillon, G. "Measuring Factors that Influence the Success of Internet Commerce," *Information Systems Research*, (13: 2), 2003, 187-204.

US Department of Commerce. "A Nation Online: Entering the Broadband Age," Report, 2004.

Table 1. Summary Statistics for DVDs

Variable	Observations	Mean	Std. Dev.	Min	Max
Product Level Characteristics					
Top 20 in location	4550615	.0405002	.1971293	0	1
Relative Price	2590829	.2529982	.0768884	.0100066	.6670557
Overall sales rank	4550615	1533.203	3011.762	1	33317
log(sales rank)	4550615	6.191366	1.612894	0	10.41382
Average rating	4548515	4.300041	.4789352	2	5
Number of Reviews	4548515	291.597	437.3478	2	3554
Days since launch	4521966	557.4311	537.4252	-40	3763
Very popular (top 25 overall)	4550615	.0371798	.1892022	0	1
Local Demand Characteristics					
Per capita income	4540314	24465.79	5693.937	9872	44962
Percent college graduate	4540314	.119227	.0361191	.0345597	.3046665
County population	4542682	1121610	1979343	5779	9937739
Percent under age 24	4540314	.3494895	.0397485	.2014658	.6605745
Broadband	4542682	5.301924	2.016319	.25	13.66667
Local Supply Characteristics					
Total specialty stores	4464531	32.41915	59.81121	0	296
Discounters	4464531	24.27556	32.90989	0	149
Big specialty stores (over 50 employees)	4464531	.3681305	.9654779	0	5
Specialty stores and discounters combined	4464531	56.69471	91.62564	0	445
Any big specialty stores	4542682	.1885241	.3911301	0	1

Table 2. Summary Statistics for Books

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Product Level Characteristics					
Top 20 in location	5534965	.1760277	1.090264	0	10
Relative Price	5534965	.2708227	.1396631	0	.6
Overall sales rank	5533486	7.019807	2.504038	0	13.63113
Average rating	5520879	4.11326	.5584671	1.5	5
Number of Reviews	5344342	4.989635	1.45267	.6931472	8.650849
Days since launch	5381083	1359.135	2178.076	-18	18523
Very popular (top 25 overall)	5534965	.0461311	.209769	0	1
Local Demand Characteristics					
Per capita income	5520745	24.42648	5.700743	9.872	44.962
Percent college graduate	5520745	.1189681	.0362002	.0345597	.3046665
County population	5523589	1102607	1958399	5779	9937739
Percent under age 24	5520745	.3490784	.0398934	.2014658	.6605745
Broadband	5523589	5.274903	2.017417	.25	13.66667
Local Supply Characteristics					
Total specialty stores	5429282	62.8548	95.14332	0	467
Discounters	5429282	23.89559	32.65378	0	149
Big specialty stores (over 50 employees)	5429282	2.385133	4.239809	0	21
Specialty stores and discounters combined	5429282	38.95921	63.52397	0	318
Any big specialty stores	5429282	0.997	0.051	0	1

Table 3. How Retailer Type Determines the Trade-off between Selection, Price, and Convenience

Retailer Size	Selection	Price	Convenience
Discounters	Small	Low	High
Big Specialty Stores	Big	High	High

Table 4. Main Hypotheses

Variable	*Expected Sign*	*Intuition*
Price		
Relative Price	Positive	As relative price increases, the product's online popularity ranking increases.
Relative Price*local discounters	Negative	If there are discounters, consumers will buy expensive items offline
Selection		
Sales Rank*Big Stores	Negative	If no big stores are present then consumers buy less popular products online.
Unpopular*Big Store Entry	Negative	If no big stores enter, then consumers buy less popular products online
Convenience		
Very Popular*Discounters	Negative	With an increase in the number of discount stores, more popular products are bought offline
Very Popular*Big Stores	Negative	With an increase in the total number of offline stores, more popular products are bought offline

Table 5. DVD Results

	Overall differences by population		Differences, controlling for discounters		Differences, controlling for large stores	
	(1)	(2)	(3)	(4)	(5)	(6)
	Population under 200k	Population over 200k	Population under 200k	Population over 200k	Population under 200k	Population over 200k
log(sales rank)	-0.004491	-0.006571	-0.000537	-0.004552	-0.004359	-0.006372
	(0.000254)**	(0.000143)**	(0.000746)	(0.000565)**	(0.000262)**	(0.000164)**
Relative price	-0.126033	-0.118167	-0.125552	-0.124618	-0.125247	-0.118369
	(0.003115)**	(0.001685)**	(0.009558)**	(0.006603)**	(0.003224)**	(0.001935)**
Very popular (top 25 overall)	-0.057646	-0.066068	-0.035250	-0.058147	-0.057920	-0.066766
	(0.001198)**	(0.000747)**	(0.003341)**	(0.002935)**	(0.001230)**	(0.000851)**
Missing price information	-0.184716	-0.182151	-0.174890	-0.189924	-0.183610	-0.181897
	(0.004067)**	(0.002194)**	(0.012496)**	(0.008598)**	(0.004210)**	(0.002520)**
Log(days since launch)	-0.007571	-0.009106	-0.007547	-0.009059	-0.007571	-0.009095
	(0.000348)**	(0.000177)**	(0.000348)**	(0.000179)**	(0.000348)**	(0.000177)**
Number of Reviews	0.000336	0.000356	0.000336	0.000359	0.000336	0.000357
	(0.000004)**	(0.000002)**	(0.000004)**	(0.000002)**	(0.000004)**	(0.000002)**
Relative price* log(ldiscounters)			-0.000213	0.002028		
			(0.005924)	(0.002035)		
Missing price* log(ldiscounters)			-0.006526	0.002450		
			(0.007739)	(0.002649)		
log(sales rank)* log(ldiscounters)			-0.002710	-0.000639		
			(0.000474)**	(0.000176)**		
Very popular* log(ldiscounters)			-0.015907	-0.002593		
			(0.002208)**	(0.000917)**		
Relative price* Any big stores					-0.012509	0.000775
					(0.012483)	(0.003941)
Missing price* Any big stores					-0.017562	-0.001288
					(0.016274)	(0.005127)
log(sales rank)* Any big stores					-0.002200	-0.000876
					(0.001077)*	(0.000336)**
Very popular* Any big stores					0.005694	0.002614
					(0.005446)	(0.001778)
Observations	1198525	3321341	1193989	3240400	1198525	3313466
Number of FEs	447252	1043916	445538	1020023	447252	1041755
R-squared	0.02	0.02	0.02	0.02	0.02	0.02
Controls	-Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Product-location fixed effects (differenced out) -Dummy for missing elapsed date information -Number of reviewers -Dummy for Walmart or Target entry within 5.4 miles, -Dummy for Barnes & Noble or Borders entry within 5.4 miles -Demographics such as broadband, percentage college educated, percentage age under 24, per capita income					

Standard errors in parentheses. All regressions include location-product fixed effects. + significant at 10%; * significant at 5%; ** significant at 1%

Table 6. DVD Results with Demographic Controls

	(1)	(1)	(3)
	Selection	Convenience	Price
log(sales rank)*	-0.00061		
Large Specialty Store	(0.000314)+		
Very popular*		-0.00306	
log(discounters)		(0.000588)**	
Relative price*			-0.01803
log(discounters)			(0.001675)**
Relative price	-0.01803	-0.2491	-0.4169
	(0.001675)**	(0.00588)**	(0.0133)**
Missing price information	0.00277	-0.0628	-0.063
	(0.000111)**	(0.000483)**	(0.000484)**
Log(days since launch)	-0.417	0.00203	0.00204
	(0.01334)**	(0.000111)**	(0.000111)**
Very popular	-0.0630	-0.0498	-0.0577
(top 25 overall)	(0.000484)**	(0.00165)**	(0.000638)**
log(sales rank)	0.002039	0.005351	0.003502
	(0.000111)**	(0.00132)**	(0.00116)**
Relative price*	0.00923	0.00893	0.0203
log(population)	(0.000433)**	(0.000434)**	(0.001358)**
log(sales rank)*	-0.00046	-0.00091	-0.00076
log(population)	(0.000097)**	(0.000116)**	(0.000088)**
log(sales rank)*	-0.05766		
log(% college graduates)	(0.003391)**		
log(sales rank)*	-0.03023		
log(% under 24 years old)	(0.00306)**		
log(sales rank)*		0.000011	
Broadband competition		(7.20e-05)	
Relative price*			0.002769
Per capita income			(0.000111)**
Observations	4,547,418	4,547,418	4,547,418
Number of FEs	1,499,314	1,499,314	1,499,314
R-squared	0.01	0.01	0.01
Controls	-Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Product-location fixed effects (differenced out) -Dummy for missing elapsed date information -Number of reviewers -Dummy for Walmart or Target entry within 5.4 miles, -Dummy for Barnes & Noble or Borders entry within 5.4 miles		

Standard errors in parentheses. All regressions include location-product fixed effects.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 7. Book Results—Difference in Difference on Store Entry

	(1)	(2)	(3)	(4)
	All Data	Towns Under 100k	Towns 100k-1 million	Towns over 1 million
(Relative Price)*	-0.0147	-0.0215	-0.0105	-0.0241
(Walmart or Target Entry within 5.4 miles)	(0.0022)**	(0.0102)*	(0.0025)**	(0.0046)**
(Relative Price)* (Barnes & Noble or	-0.0183	-0.016	-0.0271	-0.0041
Borders Entry within 5.4 miles)	(0.0061)**	-0.0196	(0.0086)**	-0.009
(Top 150 Books)*	-0.032	-0.0409	-0.032	-0.0299
(Walmart or Target Entry within 5.4 miles)	(0.0012)**	(0.0045)**	(0.0015)**	(0.0023)**
(Top 150 Books)* (Barnes & Noble or	-0.0339	-0.0067	-0.0325	-0.0389
Borders Entry within 5.4 miles)	(0.0025)**	-0.0176	(0.0029)**	(0.0045)**
(Top 150-500 Books)*	-0.0034	-0.0071	-0.0048	0.0014
(Walmart or Target Entry within 5.4 miles)	(0.0008)**	(0.0030)*	(0.0010)**	-0.0016
(Top 150-500 Books)* (Barnes & Noble or	-0.0029	-0.0025	0.0007	-0.0073
Borders Entry within 5.4 miles)	-0.002	-0.0142	-0.0024	(0.0034)*
(Top 500-1500 Books)*	-0.006	-0.0065	-0.0072	-0.0024
(Walmart or Target Entry within 5.4 miles)	(0.0006)**	(0.0028)*	(0.0007)**	(0.0013)+
(Top 500-1500 Books)* (Barnes & Noble	-0.0022	0.0112	-0.0008	-0.0061
or Borders Entry within 5.4 miles)	-0.0016	-0.0115	-0.0021	(0.0026)*
(Top 1500-5000 Books)*	-0.0082	-0.0072	-0.008	-0.0088
(Walmart or Target Entry within 5.4 miles)	(0.0009)**	-0.005	(0.0010)**	(0.0019)**
(Top 1500-5000 Books)* (Barnes & Noble	-0.0074	0.0172	-0.0068	-0.0105
or Borders Entry within 5.4 miles)	(0.0025)**	-0.0116	(0.0033)*	(0.0040)**
(Top 5000-15000 Books)*	-0.0019	-0.0006	-0.0029	0.0014
(Walmart or Target Entry within 5.4 miles)	(0.0007)**	-0.0037	(0.0007)**	-0.0016
(Top 5000-15000 Books)* (Barnes &	-0.0022	0.0167	-0.0026	-0.0049
Noble or Borders Entry within 5.4 miles)	-0.0018	-0.0157	-0.0025	(0.0024)*
Relative price	0.0237	0.0221	0.0221	0.0275
	(0.0007)**	(0.0023)**	(0.0009)**	(0.0015)**
Top 150 Books	0.006	0.005	0.0059	0.0066
	(0.0004)**	(0.0014)**	(0.0005)**	(0.0008)**
Top 150-500 Books	-0.0124	-0.0104	-0.0122	-0.0135
	(0.0003)**	(0.0010)**	(0.0004)**	(0.0006)**
Top 500-1500 Books	-0.0061	-0.0041	-0.006	-0.0069
	(0.0003)**	(0.0009)**	(0.0003)**	(0.0005)**
Top 1500-5000 Books	-0.0066	-0.0047	-0.0063	-0.0077
	(0.0002)**	(0.0007)**	(0.0003)**	(0.0004)**
Top 5000-15000 Books	-0.0041	-0.0023	-0.0041	-0.0046
	(0.0001)**	(0.0004)**	(0.0002)**	(0.0002)**
Observations	4062326	386551	2556060	1119715
Number of FEs	981255	93393	616443	271419
R-squared	0.07	0.07	0.07	0.08
Controls	-Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Product-location fixed effects (differenced out) -Dummy for missing elapsed date information -Number of reviewers -Dummy for Walmart or Target entry within 5.4 miles, -Dummy for Barnes & Noble or Borders entry within 5.4			

	miles -Demographics such as Broadband, percentage college educated, percentage age under 24, per capita income

Standard errors in parentheses and are clustered by location-time. All regressions include location-product fixed effects. + significant at 10%; * significant at 5%; ** significant at 1%

Figure 1a. Characteristics of Top Selling DVDs by City Size

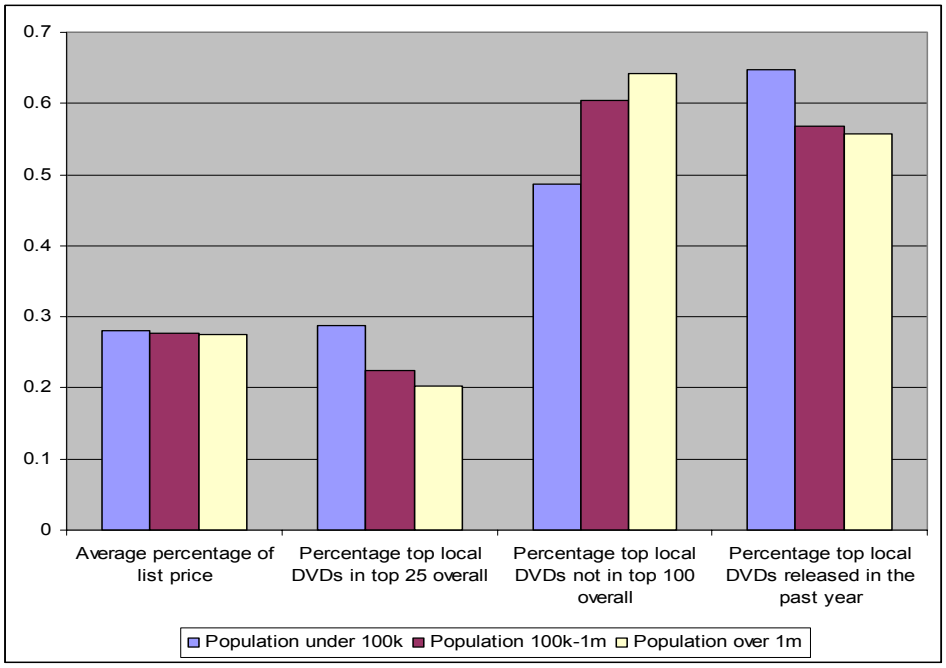


Figure 1b. Characteristics of Top Selling Books by City Size

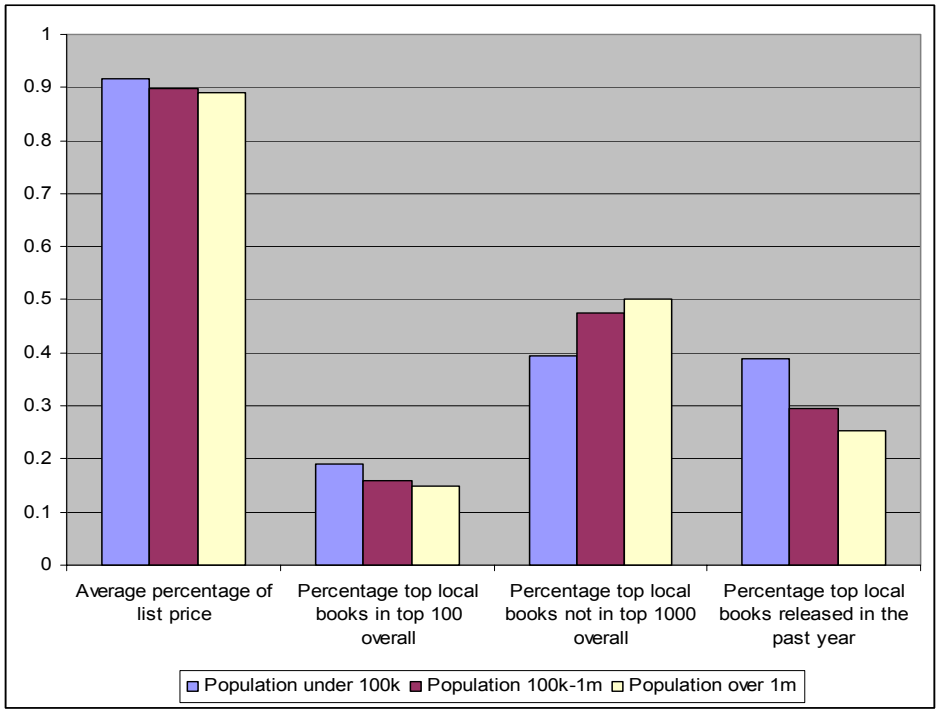
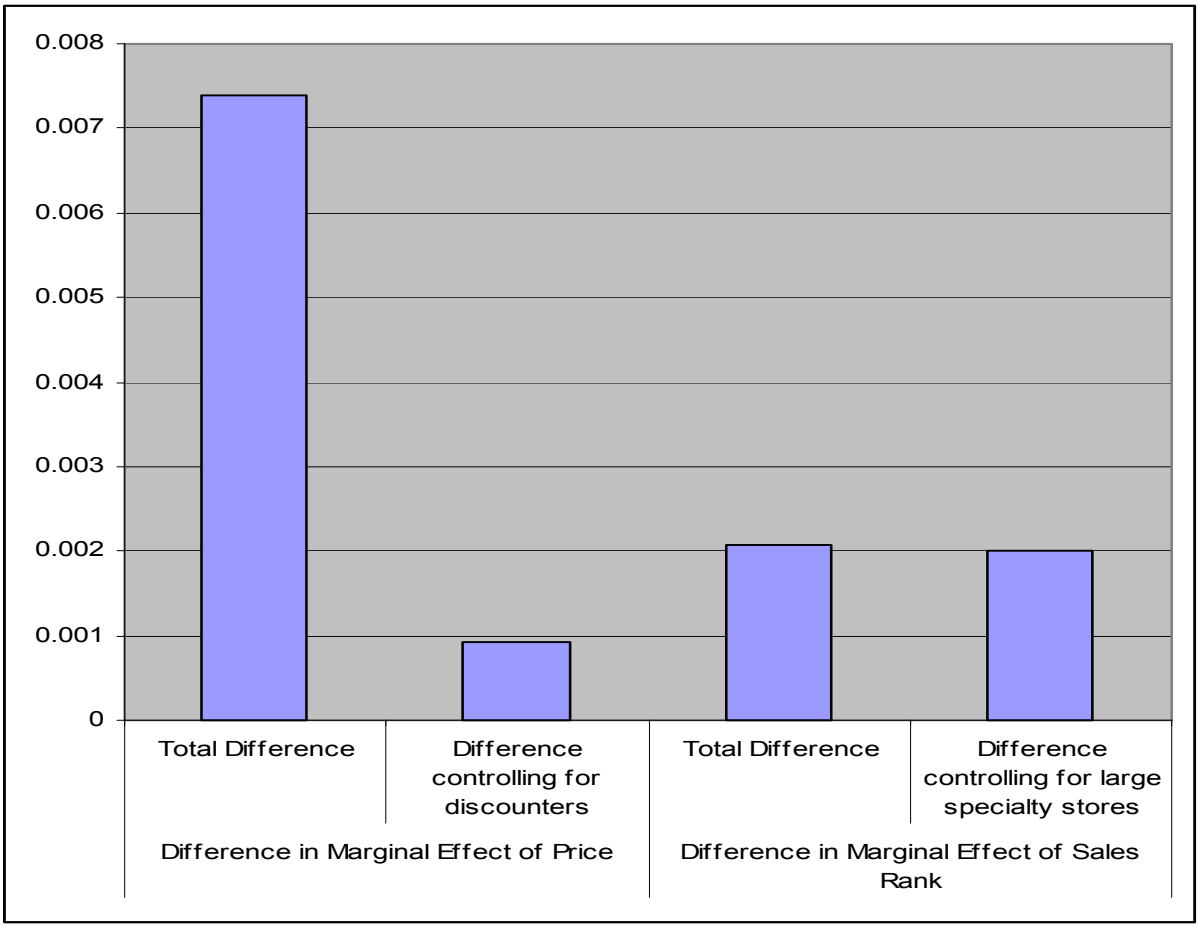


Figure 2. Amount of Difference between Urban and Rural Areas Explained by Discounters and Large Specialty Stores



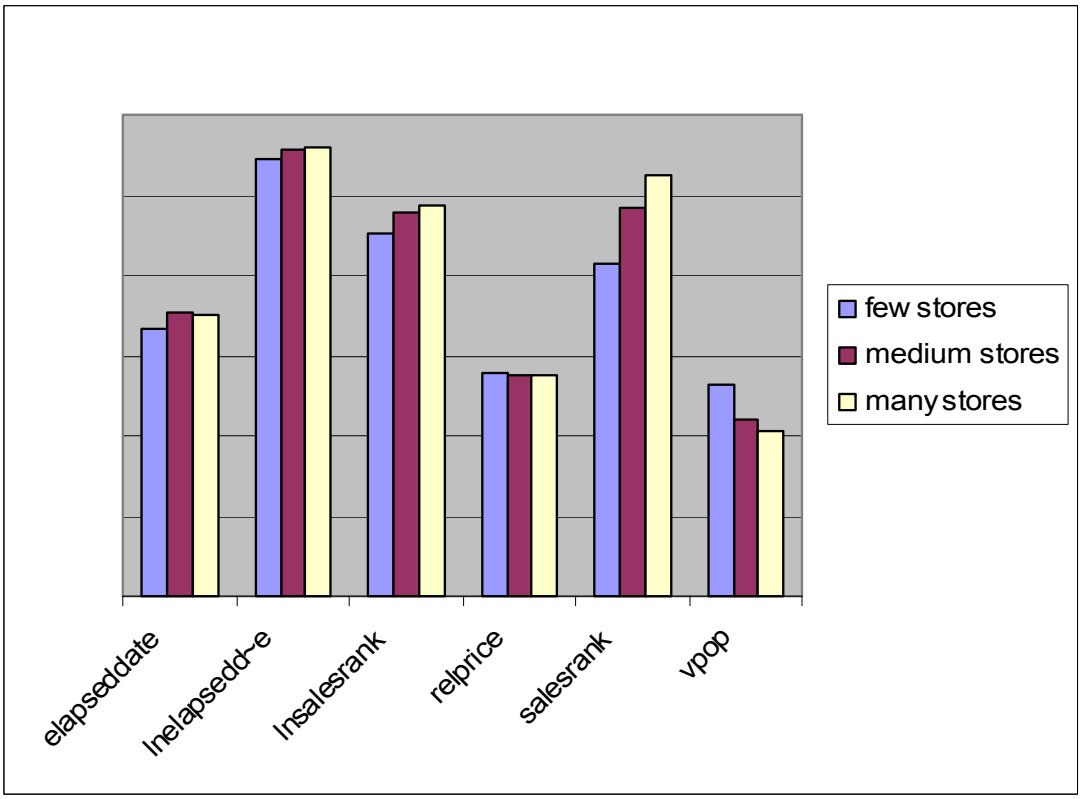


Figure 3a. How the Presence of Discounters Affects the Ooverall Choice of Products.

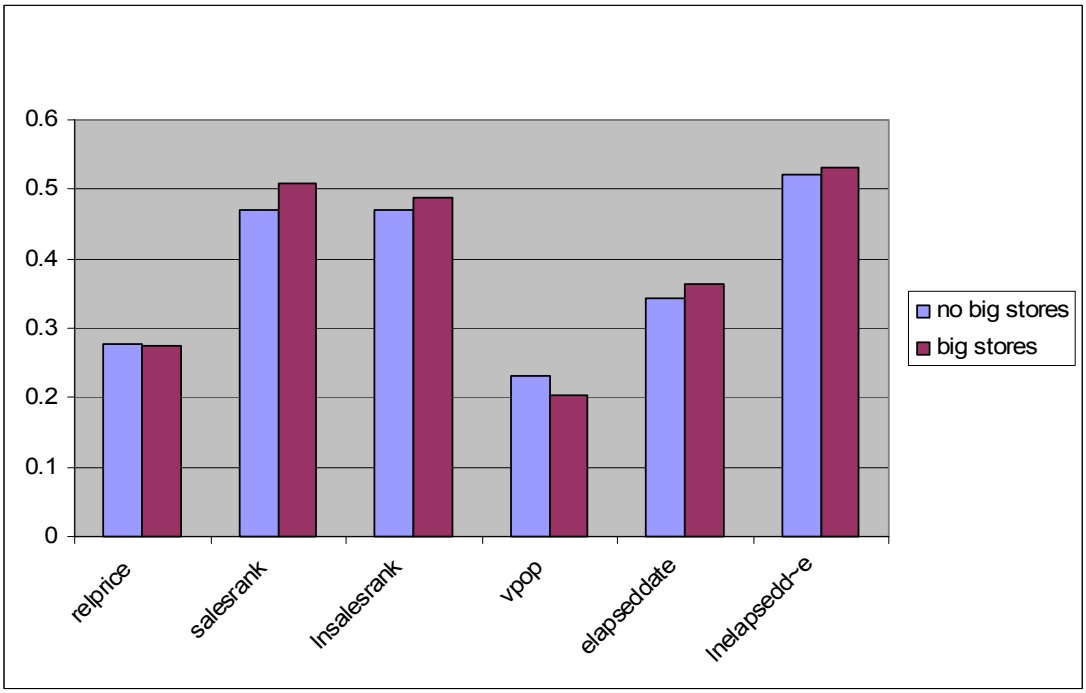


Figure 3b. How the Presence of Big Stores Affects the Overall Choice of Products.

Figure 4a. Intuition for Proposition 1 *Online Purchasing for Convenience*

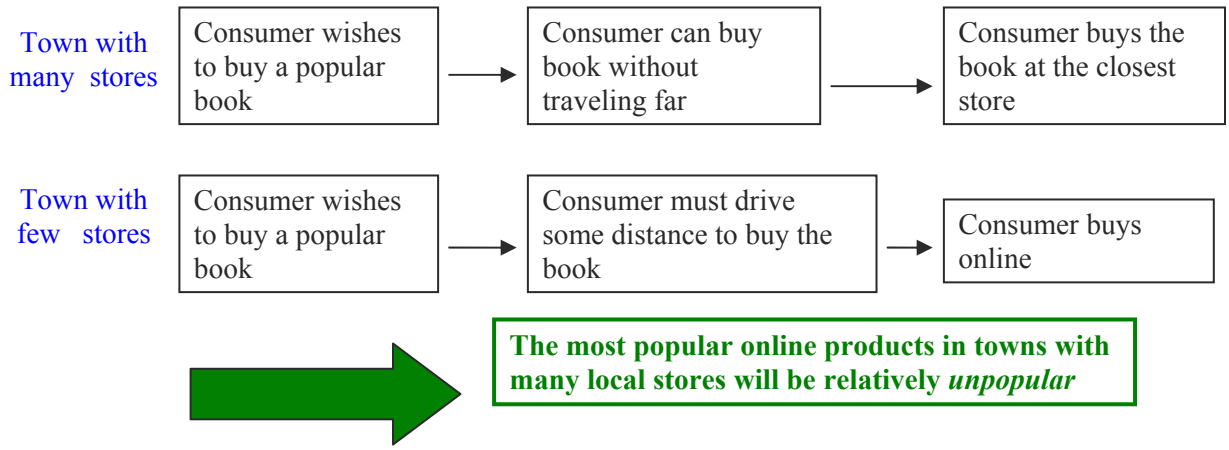


Figure 4b. Intuition for Proposition 2 *Online Purchasing for Selection*

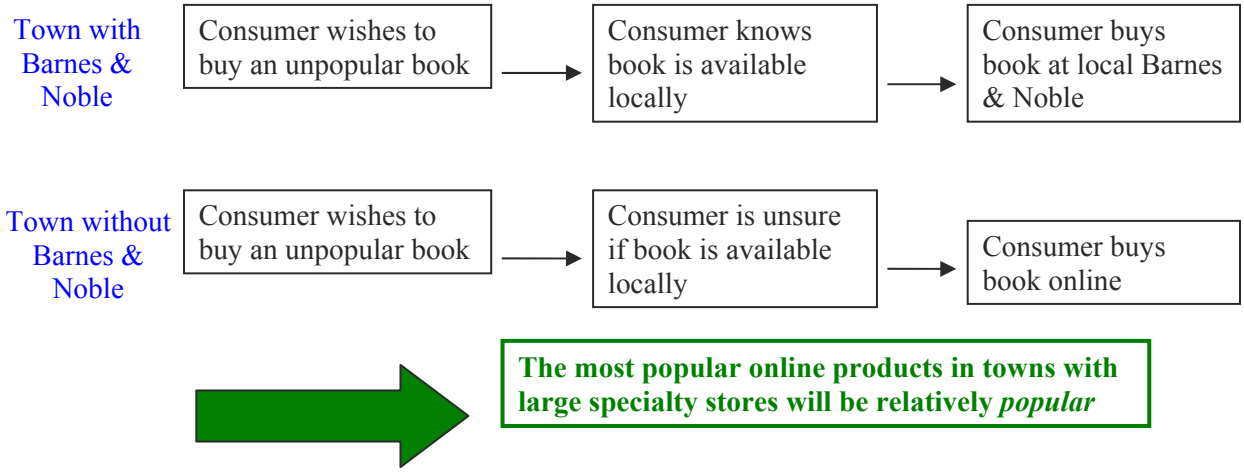


Figure 4c. Intuition for Proposition 3 *Online Purchasing for Low Prices*