Geography and Electronic Commerce: Measuring convenience, selection, and price[*]

Chris Forman
Tepper School of Business
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
cforman@andrew.cmu.edu

Anindya Ghose
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
aghose@stern.nyu.edu

Avi Goldfarb
Rotman School of Management
University of Toronto
105 St George St
Toronto, ON M5S 3E6
agoldfarb@rotman.utoronto.ca

February 2007

Abstract

We examine how the local availability of offline retail options drives use of the online channel and consequently how the convenience, selection, and price advantages of the online channel may vary by geographic location. In particular, we examine the effect of local store openings on online book purchases in that location. We explore this problem using data from Amazon on the top selling books for 1501 unique locations in the US for 10 months ending in January 2006. In addition to this data, we use information on changes in local retail competition as measured by openings of large bookstores such as Borders or Barnes & Noble and discount stores such as Wal-Mart or Target. We show that even controlling for product-specific preferences by location, changes in local retail options have substantial effects on online purchases. We demonstrate how the convenience, selection, and price benefits of the Internet are different for consumers in different types of locations. More generally, we show that geography significantly impacts the benefit that consumers derive from electronic markets.

[*] We thank participants at the Second Annual Symposium on Statistical Challenges in E-Commerce, the Marketing Science 2006 Conference, the INFORMS 2006 Conference, the 2006 International Conference in Information Systems, the 2006 Harvard-Wharton Consortium on Operational Excellence in Retailing, the HICSS 20th Anniversary Symposium on Competitive Strategy, Economics, and Information Systems, the Federal Trade Commission, DIS Workshop in University of Florida, seminar participants at Carnegie Mellon University, Georgia Tech, University of Toronto, Michigan State University, University of Texas Austin, New York University, University of South Carolina, and three anonymous reviewers of the International Conference in Information Systems (ICIS 2006) for useful feedback. We also thank Jeremy Fox, Korhan Gurkan, Ke-Wei Huang, Steven Klepper, Barrie Nault, Paul Pavlou, and Jeffrey Prince for valuable suggestions. We thank Rong Zheng and Ashley Tyrrel for outstanding assistance with the collection of data, and Gordon Eiland, VP of Strategy at Borders for supplying us with the store opening data of Borders. Support was provided by a Berkman Faculty Development grant, the NET Institute, NYU Research Challenge grant # N6011 and SSHRC grant # 538-02-1013. All errors are ours alone.

1. 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. Prior work has identified how the existence of electronic markets improves consumer welfare by providing lower prices (e.g. Brynjolfsson and Smith 2000; Clemons, Hann, and Hitt 2002), offering greater selection to consumers (e.g., Brynjolfsson, Hu, and Smith 2003; Ghose, Smith and Telang 2006) and providing greater convenience by eliminating travel costs and enabling 24x7 purchases irrespective of geographic location (Cairncross 1997). In sum, an evolving body of work has demonstrated that electronic commerce has provided consumers with better selection, greater convenience, and lower prices.[1]

While the benefits of electronic commerce have been explored at some length, there is relatively little understanding of how consumers substitute between online and offline channels. Existing work has demonstrated that consumers will shift from offline to online channels to obtain better prices (Chiou 2005; Goolsbee 2000; Prince 2006). However, we have little understanding of how consumers substitute between online and offline channels to benefit from increased convenience and selection. Obtaining a better understanding of online-offline channel substitution is essential for both traditional and Internet retailers. Local retail options may be an important determinant of online choices. For example, if consumers use Internet channels primarily to obtain lower prices for or more convenient access to very popular items, then the diffusion of large discount retailers such as Wal-Mart into new locations will result in a long run shift in buying patterns away from the most popular products at online retailers.

We develop a set of hypotheses motivated by the existing theory literature on channel competition for commodity products. Most of the literature we emphasize adjusts the circular city model of Salop (1979) to include a direct marketer (Balasubramanian 1998, Cheng and Nault 2005, Jeffers and Nault 2005, Viswanathan 2005). In these models, a consumer's choice to use the online or offline channel is a function of the prices in each channel, the transportation cost and distance to the closest offline

[1] Electronic commerce may benefit consumers or influence purchase behavior in other ways such as by providing consumers with additional information to make better product decisions through product reviews (e.g., Dellarocas 2003). Though it is not our primary focus, we do include this as a control in our analysis.

retailer, and a measure of the suitability of the product for the online channel. We use this literature to motivate three testable hypotheses that each relate to one of convenience, selection, and price. We then develop additional testable predictions from changes in local market characteristics and in the suitability of the product for the online channel. To the extent that our hypotheses are supported, our results provide some validation for many of the modeling assumptions in spatial differentiation models.

We test these hypotheses by examining changes over time in online purchases in 1501 local geographic markets in the US. We control for differences in consumer preferences across locations through product-location fixed effects, and use local retail store entry to identify the effects of improved offline options on online choice using a difference-in-difference methodology. In particular, we utilize changes in consumer behavior after local retailer entry to identify whether consumers utilize online channels primarily to obtain improved convenience, selection, or price.

Our data come from the "Purchase Circles" web pages on the Amazon.com web site. The pages include information on the top-selling books in local geographic markets throughout the US. We use monthly data from April 2005 to January 2006. Although these data focus on the top-selling books in each market, there is considerable variance in best sellers across markets. Moreover, as Amazon.com is by far the largest online retailer, these data provide an excellent setting for measuring differences in the use of online channels across US locations.

We find strong evidence for *convenience* across all location types: after either a discount retailer (Wal-Mart or Target) or a large specialty store (Barnes & Noble or Borders) enters a market, we find that local online purchases of the nationally most popular products decline. This suggests that consumers are more likely to purchase popular products offline in places where offline shopping is more convenient. We also find evidence that consumers substitute from online channels to obtain lower *prices*. These results are particularly strong in settings where the attractiveness of online commerce is relative low, such as when consumers reside in states with sales taxes or when purchasing products in which the online channel is relatively less effective. Overall, we do not find consistent evidence of selection effects in online-offline channel substitution. We cannot say whether the lack of an overall selection effect is due to insufficient

power of our test or because there actually is no selection. However, we do find evidence of *selection* in university towns and larger cities, where the concentration of consumers with preferences for less popular products is greater. More generally, we find that geography matters: the actions of local offline retailers affect online retailers of commodity products.

1.1 Related Literature and Contribution

This paper advances several streams of research. First, it examines the empirical relevance of recent theoretical models of online-offline channel substitution. In particular, our paper is most closely related to theoretical models of spatially differentiated commodity markets that are derived from Salop's (1979) circular city model (Balasubramanian 1998, Cheng and Nault 2005, Jeffers and Nault 2005, Viswanathan 2005). As we will show below, several of our hypotheses are direct implications of these models, and to our knowledge we are the first authors to provide empirical validation of the assumptions of these models. This paper is also related to models of channel substitution that allow for differences in tangible and intangible components of the bundle of products and services offered in the two distribution channels (Lal and Sarvary 1999, Riggins 2004, Druehl and Porteus 2005).

Our paper also contributes to the empirical literature on consumer substitution between online and offline channels. Existing work focuses on the cross-price elasticity between online and offline channels. For example, both Goolsbee (2000) and Ellison and Ellison (2006) show that tax avoidance may be an important factor in e-retail activity. Prince (2006) measures changes in the cross-price elasticity of personal computer sales across channels. In contrast, our research focuses on how changes in offline price, convenience, and selection influence online consumer decisions. To our knowledge, no prior work in this area has examined how channel substitution is influenced by the convenience and product selection of the offline option.

More broadly, this paper advances the emerging stream of literature that studies how Internet retailing contributes to consumer welfare. One stream of this literature has studied how Internet retailing has influenced price competition and price dispersion (e.g. Brynjolfsson and Smith 2000), and

demonstrated that consumers benefit from lower prices in the online channel.[2] A related line of research has shown that by lowering search costs, Internet retailing improves consumer welfare in other ways. Lower search costs help consumers obtain hard-to-find books (Brynjolfsson, Hu, and Smith 2003), increase the resale value of new products (Ghose, Telang and Krishnan 2005), and facilitate the market for used books (Ghose, Smith, and Telang 2006).

We contribute to this broad literature in two ways. First, we emphasize the benefits of Internet retailing in improving customer convenience. Though prior research has examined how Internet technology reduces the costs associated with distance (Forman, Goldfarb, and Greenstein 2005; Sinai and Waldfogel 2004) and improves the information convenience in online trading markets (Balasubramanian, Konana and Menon 2003), the importance of the Internet channel in improving convenience by reducing transportation costs has thus far played a relatively minor role in electronic commerce research. Second, we provide a framework for measuring the relative importance of price, selection, and convenience on consumers' channel decision. In contrast, prior work has only been able to study these individual mechanisms in isolation.

We show that the online channel means different things to people in different locations. The Internet was touted to make geographic location irrelevant. However, we show that geography matters, even online. A number of recent papers have examined how consumer behavior varies across geographic markets by incorporating spatial structure into their models (Bronnenberg and Mahajan 2001; Jank and Kannan 2006). These papers suggest that spatial data captures both demand and supply factors that vary with location. We focus on local retail supply and show how it influences online behavior.

Our results have significant managerial implications for online and offline retailers. For online retailers, our research shows how consumers' use of the online channel varies across locations. Depending on local retail options, the importance of price, convenience, and selection varies. Understanding this helps frame promotional strategies and product offerings in online channels to different locations. For offline retailers, our work shows how consumers use the online channel to

[2] See Baye, Morgan, and Scholten (2006) for a review of research in this area.

substitute for their offerings and hence provides a better understanding of the potential threat from the online channel. Conventional wisdom suggests that geographic differentiation is an important factor allowing offline retail stores to maintain the markups over marginal cost they need to survive (Ellison and Ellison 2006). This benefit of geographic differentiation may be limited when the competition is from the online world.

The rest of the paper is as follows. In Section 2, we draw on the existing theoretical literature to generate our hypotheses. Section 3 and 4 describe the data and empirical model respectively. Section 5 provides the analysis along with some robustness checks and extensions. We conclude in Section 6 with some discussion of managerial and research implications.

2. Theory and Hypotheses

Our research examines how spatial location influences the behavior of consumers online. To this end, our empirical analysis is motivated by the theoretical literature on spatial competition, especially a number of papers based on Salop's (1979) circular city model. In Salop's model, consumers are uniformly distributed around a circle and each consumer is in the market for one unit of a commodity product. Consumers incur a transportation cost when visiting a retailer: these costs can include the monetary costs of travel, as well as inconvenience costs and the opportunity cost of time.

Balasubramanian (1998) extends this model of Salop by adding a direct marketer (or online retailer). Consumers who purchase from the direct marketer do not incur a transportation cost, but they do incur a disutility cost μ that can be thought of as the product's lack of fit with the direct channel. For example, it can capture the costs of delayed gratification, or product quality uncertainty. Balasubramanian uses this model to examine how entry by the direct marketer influences outcomes in the offline market.

We use results from the circular city models of Balasubramanian (1998) and related studies by Cheng and Nault (2005), Jeffers and Nault (2005), and Viswanathan (2005) to motivate our hypotheses.[3] However, our empirical setting differs from these models in one significant way. In our setting consumers

[3] For other examples of use of the Salop model in the other contexts of the Information Systems literature, see Bakos (1997) and Dewan, Freimer and Seidman (2000).

demand a large set of commodity products that vary in popularity. In contrast, most circular city models focus on a single commodity product. While Internet retailers can effectively stock even the most unpopular products at low cost, offline retailers must decide on finite inventory levels that constrain their ability to offer unpopular products. For example, small stores stock approximately 20,000 unique titles, while large independent booksellers stock between 40,000 and 100,000 unique titles (Brynjolfsson, Hu and Smith 2003).

For concreteness sake, we will refer to three types of products below: *popular products* that have the greatest sales in a particular category and which are likely to be stocked by any retailer; *less popular products* that have fewer sales and require significant investments in shelf and inventory space for the retailer to stock; and *unpopular products* that have so few sales they are unlikely to be economically stocked by any offline retailer, these are products in the "long tail" of the product sales distribution (e.g., Anderson 2006). This distinction is important in light of our data. We have information on the *relative* online sales of these products by consumers in a number of cities and towns. We do not have total sales figures. Therefore our hypotheses focus on how the relative sales of popular products, less popular products, and unpopular products vary across locations.

2.1. Main hypotheses: Price, convenience, and selection

According to Balasubramanian (1998, p. 183), "Convenience is a central reason for the patronage of direct channels." Our first hypothesis describes how increased convenience influences consumers' behavior in the online channel. We formulate and test this hypothesis by examining how distance to a retailer shapes the propensity of a consumer to purchase popular products online. Balasubramanian shows that in a single-good commodity market in which there are no inventory constraints, increases in the distance to the closest offline retailer will increase the likelihood that a consumer will purchase from the direct (Internet) retailer, other things equal. This is a direct implication of the presence of travel costs that are increasing in the distance to the closest offline retailer.

In our setting, popular products are likely to be stocked by all offline retailers. Thus, the

likelihood that a consumer will purchase popular products online will be increasing in the distance to offline stores. The likelihood that consumers will purchase less popular and unpopular products online will be increasing more slowly in distance than that of popular products, because these products are less likely to be available at offline retailers. This gives rise to our first hypothesis:

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

Our second hypothesis describes how broader selection influences online behavior. As the product selection at offline retailers increases, consumers will be able to increasingly purchase less popular products offline. Thus, for every consumer on the unit circle, the likelihood of purchasing a less popular product online will decrease as offline product selection increases, because the consumer will now be able to purchase these products offline. Intuitively, locations with a Barnes & Noble will have a wider selection of books than locations with just a Wal-Mart. The effect of a Barnes & Noble on online purchases will then be different from the effect of a Wal-Mart.

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

The next two related hypotheses are statements about how consumers respond to changes in online and offline price. The first such hypothesis simply says commodity products that are discounted will be bought more frequently than products that are not discounted. While we believe the focus on relative sales makes this hypothesis worth articulating, it is essentially a statement that demand curves are negatively sloped.

Hypothesis 3a: Online purchasing for price: *As the discounted online price of a commodity falls relative to its list price, consumers are more likely to buy it. This increases the relative online sales of that commodity compared to other online items.*

Hypothesis 3b says that the role of discounts in determining the relative sales of different online commodities (as in Hypothesis 3a) will depend on the distance to offline stores. Items that are discounted

online are also typically discounted offline. For example, best sellers are typically discounted both online and offline. Therefore, the impact of online discounts will be tempered by the existence of local retail stores. This is a direct implication of the circular city model of Balasubramanian (1998). Decreases in the distance to offline retailers will increase the utility of the offline option, making consumers less sensitive to price changes online.

Hypothesis 3b: Online purchasing for relative price: *As distance to offline stores falls, discounted online items will be more likely to be bought offline.*

2.2. Comparative statics on the effects of price, convenience, and selection

The next set of hypotheses examines how the relationships defined in hypotheses 1-3 are influenced by other market and product characteristics.

The first such hypothesis says that although increases in distance will lead to a greater fraction of popular products purchased online, this relationship will be increasing in the distance to nearby retail options. This hypothesis is motivated by several streams of prior research. Bresnahan and Reiss (1991) show that the impact of the marginal entrant on local market competition is declining in the number of existing competitors. Cheng and Nault (2005) show the importance of market coverage in a circular city model. In uncovered market segments, offline retailer entry will have a greater impact on channel substitution than in covered market segments. In our setting, the testable implication of this research is that the effect of the marginal offline retailer on a consumer's decision to purchase offline is declining in the number of offline retailers. We use local market size to proxy for the number of retailers and market coverage. Therefore, we expect

Hypothesis 4a: Market size and convenience: *The relationship between the distance to offline stores and the proportion of popular books purchased online (stated in Hypothesis 1) will be stronger in small markets than in large markets.*

Consumer preferences may vary within and across markets. In small markets with homogeneous tastes, sales of less popular products are likely to be low, both online and offline. In large markets, or

markets with heterogeneous tastes for books (such as university towns), less popular products will constitute a higher proportion of sales. Waldfogel (2003) showed that larger, heterogeneous markets have more media variety due to diversity in consumer preferences. Applying this result to book sales, the advantages of selection examined in hypothesis 2 should matter more in locations with a sufficient number of consumers who are interested in less popular products.

Hypothesis 4b: Market size and selection: *The relationship between availability in offline markets and the proportion of less popular products bought online (stated in Hypothesis 2) will be stronger in heterogeneous markets such as those in which universities are located and in large markets.*

Our second comparative static hypothesis relates to taxes. Sales taxes are assessed for Amazon purchases in only the four states in which Amazon has distribution centers: Kansas, Kentucky, North Dakota, and Washington. The utility of purchasing commodity products from the online channel is lower for consumers located in these states, other things equal. In the language of Balasubramanian's (1998) adaptation of the circular city model, in these locations the suitability of the online channel (μ) is relatively poor compared to states with no online sales taxes. Prior work has demonstrated that the assessment of online sales taxes influences consumer propensity to buy online (Goolsbee 2000; Ellison and Ellison 2006). Thus, it is not our primary goal to assess the impact of sales taxes on electronic commerce. Rather, we seek to examine how online sales taxes influence the relationship between distance to offline retailers and what consumers purchase online. In particular, we expect that because the cost of purchasing online is higher for consumers in sales tax states, the estimates of our parameters measuring the effects of convenience, selection, and price will be larger (in absolute value) in states with sales tax than in others. Intuitively, competition from offline stores will be fiercer in sales tax states.

Hypothesis 5: Taxes and price, convenience, and selection: *The magnitude of substitution behavior predicted by our price, convenience, and selection relationships (Hypotheses 1 through 3) will be larger in locations with online sales taxes.*

Our final hypothesis relates to the fit of the product with the online channel. This is therefore

another examination of the impact of changing the parameter μ in Balasubramanian's model. The online channel has several benefits over the offline channel, but it also has several drawbacks. Lal and Sarvary (1999) note that products have digital and non-digital attributes, and while Internet retailers may have a comparative advantage in presenting digital attributes, they likely have a comparative disadvantage in presenting non-digital attributes. In particular, the offline channel is relatively well-suited to undirected browsing that may lead to purchase because of the ability to visually inspect items (Wolfinbarger and Gilly 2001). Consequently, items with higher uncertainty in consumption utility may be more suited to the offline channel because of the ability to browse. In our setting, the expected utility to an individual from reading a particular fiction book may be less certain than the expected utility from reading a particular non-fiction book. Thus, as was the case for states with online sales taxes, we expect that the estimates of our parameters measuring the effects of convenience, selection, and price will be larger (in absolute value) for fiction books than for other types of products.

Hypothesis 6: Fiction books and price, convenience, and selection: *The magnitude of substitution behavior predicted by our price, convenience, and selection relationships (Hypotheses 1 through 3) will be larger for fiction books than for nonfiction books.*

3. Data Description

To examine how online behavior varies with offline supply conditions, we require detailed data on how consumer purchases vary across local geographic markets. The data that we use come from the web pages on "Purchase Circles" from the Amazon.com web site. Amazon's Purchase Circles are specialized best-seller lists that denote the top-selling books by location throughout the US.[4] We used a JAVA spider to extract and parse data from Amazon's website. Between April 2005 and January 2006 we collected monthly data on purchases for each location in the Purchase Circles. An observation in our data consists of a particular product-location-time. Some locations in our Purchase Circles raw data set do not appear for the entire time period. This will matter if locations only appear in Purchase Circles when there

[4] Henceforth, we use the word *locations* to refer to small and large cities, as well as small towns.

are a sufficient number of purchases at Amazon because it is possible that local entry by retails stores may influence whether a location appears at all. To address this concern, we use only those locations that appear in the sample throughout the 10 months.[5] This resulted in 1501 locations.

While previous studies have used data from Amazon, our use of the Purchase Circles data to understand channel substitution is unique. To our knowledge the only other study to use the data available through Purchase Circles is Forman, Ghose, and Wiesenfeld (2006). However, their use of the data is very different from ours. They use Purchase Circles to study the relationship between product reviews and sales. Bajari, Fox, and Ryan (2006) use similar data from Amazon that ranks mobile phone carrier share in 22 large US markets to examine market power.

For each location, Amazon provides a list of the top 10 selling books. Our primary dependent variable, $LocalTop10_{ijt}$, is a binary variable that is equal to one if book i is present in the local top 10 in location j in month t, and zero otherwise. Though our data contain only information on the books that appear in the top 10 in a location, there is considerable heterogeneity in this measure across locations and over time. Consumers buy different books in different locations. Figure 1 shows that in May 2005, 58.6% of products in our sample appear in the top 10 products at five or fewer locations, while only 1.5% of products appear more than 1000 times.

In the next several paragraphs, we provide further information on the construction of our independent variables. Descriptive statistics are provided in Table 1.

Product Characteristics: We use information on product details from Amazon's web site. For any given book that is listed in Purchase Circles, we collected data on the specific characteristics of the book, based on its ISBN number from Amazon's corresponding product details page. These include the offline (list) price, Amazon's retail price, the product's national sales rank on Amazon, the release date of the product in the market, the average rating from Amazon's customers, and the number of reviews posted on

[5] Due to a managerial decision at Amazon related to the threshold for inclusion in Purchase Circles, the number of locations in the sample expanded significantly in November 2005. We only include locations that are observed before and after this date.

Amazon. Shipping costs are identical across locations and are therefore not included.

To measure the price benefits of online retailing, we construct another variable that we label *Relative Price.* The *Relative Price* variable is computed as the difference between the Amazon retail price and the undiscounted offline price, normalized by the offline price. In addition to price, we examine the national rank (popularity) of a book on Amazon.[6] To allow for a flexible functional form, we compute a series of dummy variables (a spline) that indicate the specific range of national sales rank for which the book appears in that month: top 150, 151-500, 501-1500, 1501-5000, 5001-15,000, or greater than 15,000 (which we use as the base). As a robustness check, we have also used a log-linear continuous variable and the results are qualitatively similar. We define popular books as those that fall in the top 150 nationally and those that fall in the range 151-500. Books with national sales ranks in the lower ranges, specifically those not in the top 5000, are classified as less popular books.

To construct our final data set, for each month we identified the 300 books that were most frequently listed in the local top 10 lists. To include products that were listed in the top 10 in some locations but were not in this group of 300, we constructed an "outside option" product. This outside option product had characteristics equal to the average of products in this set. Our results are robust to the exclusion of the outside option. The main results are also robust to including the 1000 books that appeared most frequently in the top 10 lists. We chose 300 because the product-location fixed effects made little sense for books that only were in the top 10 in any location once or twice.

Store Entry: Our main analysis examines how offline retail store entry influences buyer choice online. Retail store entry in a given location will decrease the average distance consumers in that location must travel to get access to offline retailers, and also increase the availability of any given product, other things equal. To identify price, selection, and convenience effects, we examine entry of two types of stores. For each location in our data set, the variable labeled *Discount Store Entry* is equal to one for every month

[6] We use the sales rank on Amazon to define the national popularity of a book, both online and offline. Our results on popular books are robust to using USA Today's rank of the top 150 books in the US. We focus on the Amazon rank because it provides rankings for all books, not just the most popular.

after a Wal-Mart or Target store has entered within a 5.4 mile radius of the location; our variable labeled *Large Bookstore Entry* is equal to one for every month after a Barnes & Noble or Borders bookstore has entered within a 5.4 mile radius of the location. These data were collected either through press releases from the companies or through direct communication with company representatives. To compute radii, we use the average longitude and latitude across zip codes within the location. We use 5.4 miles because this is the distance that the average consumer travels to go to a bookstore (Brynjolfsson and Smith 2000), although we show that the results are robust (and in fact stronger) when we use a larger radius of 20 miles. Across our entire sample, 16.4% of locations experience discount store entry, while 4.7% experience a large bookstore entry.

4. Econometric Model

Our estimation strategy is based on difference-in-differences identification. We examine how the types of products bought in a location change after offline store entry, relative to a location that does not experience such entry. To be clear, in section 2 we developed a series of hypotheses based on cross-sectional variation in distance to stores across locations. Empirically, we convert these into hypotheses based on changes in local retail supply over time within locations. Focusing on changes over time is essential for identification. Unlike cross-sectional estimation, the 'within' estimates of the difference-in-difference allow us to control for local tastes and other fixed local characteristics in examining the impact of local retail conditions.

In particular, we estimate a linear probability model of whether a product *i* is in the top 10 in location *j* in month *t*.[7] We observe only a discrete measure of rank: whether a product is in the top 10 in a location. The use of rank data, rather than quantity data, means that our empirical framework will be different than those typically used to examine channel substitution. In particular, it means our analysis must be based on relative rather than absolute sales. In particular, we estimate the following linear

[7] Since our results are primarily based on interaction terms, using a non-linear model such as Probit model would have been difficult to interpret because the cross-partial of a non-linear model may have a different sign than the coefficient on the interaction term (Ai and Norton 2003). The main disadvantage of using a linear model is reduced efficiency. Given the large number of observations, this becomes less important.

regression:

$$(LocalTop10_{ijt}) = \alpha_0 + \alpha_1 DiscountStoreEntry_{jt} + \alpha_2 LargeStoreEntry_{jt} + \boldsymbol{\beta NationalRank}_{it} + \boldsymbol{\gamma NationalRank}_{it} \; x \; DiscountStoreEntry_{jt} + \boldsymbol{\delta NationalRank}_{it} \; x \; LargeStoreEntry_{jt} + \theta_1 RelativePrice_{it} + \theta_2 RelativePrice_{it} \; x \; DiscountStoreEntry_{jt} + \theta_3 RelativePrice_{it} \; x \; LargeStoreEntry_{jt} + \phi X_{it} + \mu_{ij} + \mu_t + \varepsilon_{ijt}$$

where $(LocalTop10_{ijt})$ is a dummy variable for whether product i is in the top 10 in location j for month t; $DiscountStoreEntry_{jt}$ and $LargeStoreEntry_{jt}$ indicate whether a discount store or large bookstore entered location j prior to time t; $\boldsymbol{NationalRank}_{it}$ is a vector of dummy variables for the national sales rank of product i at time t defined above; $RelativePrice_{it}$ is the relative price variable defined above; X_{it} are other attributes of product i for month t;[8] μ_{ij} is a product-location fixed effect, μ_t is a month fixed effect, and ε_{ijt} is a product-location-month idiosyncratic error term. The product-location fixed effect, μ_{ij}, controls for the overall preferences of each location for each product. We estimate this regression by differencing the average values across product-location. Standard errors are clustered over location-months.[9]

Our broad hypotheses from Section 2 easily convert into testable hypotheses on the coefficients on the interaction of local supply characteristics and product characteristics. Hypothesis 1 on convenience suggests that decreases in distance to offline stores will be associated with relatively fewer purchases of popular products online. Entry by discount stores and large stores will decrease such distances, other things equal. Therefore the coefficients on the interactions between *DiscountStoreEntry* or *LargeStoreEntry* with our *NationalRank* dummies for books that are nationally in the top 500 will be negative (i.e. the corresponding elements of γ and δ will be negative for books nationally in the top 500). Hypothesis 2 on selection suggests that entry by large bookstores with larger selection will be associated with fewer purchases of less popular products online relative to entry by discount stores. Therefore, to identify the effects of selection (Hypothesis 2), we examine two things. First, we examine whether Barnes & Noble and Borders entry is negatively associated with purchases of less popular books (those nationally

[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.

[9] We also experimented with clustering over product-months and the results are qualitatively similar. Also, it is possible that the error differs by location size because the local popularity ranking could have a different random component in smaller locations. This would lead to measurement error in the dependent variable, thereby adding heteroskedasticity to the error term. We control for this using robust standard errors.

ranked 5001 to 15,000). Second, we examine whether Barnes & Noble and Borders entry has a more negative impact on these less popular books than Wal-Mart and Target entry. Hypothesis 3a suggests that as online prices fall relative to offline prices, online sales rise: i.e., $\theta_1 < 0$. Hypothesis 3b suggests that entry by discount stores and large stores will mitigate this effect because they discount the same types of books as the online retailer: i.e., $\theta_2 > 0$ and $\theta_3 > 0$. Hypotheses 4 through 6 are identified off of differences in the magnitude of these coefficient estimates for different locations and products: small market versus large markets; university towns versus those without a university; locations with online sales taxes versus those without; and fiction versus nonfiction books. The expected signs of the coefficients based on these hypotheses, and our results, are summarized in Table 4.

Identification of our hypotheses is obtained from changes in the number of retailers within a location over time. Our product-location fixed effects control for differences in consumer preferences across locations. Moreover, by examining store entry rather than cross-sectional differences in supply, we ensure that cross-sectional differences in location-specific prices and unobserved preferences for popular books do not drive results. Instead, results will be driven by variation in local supply characteristics. Our model does require the identification assumption that store entry is uncorrelated with changes in preferences toward less popular books or less price sensitivity. However, our data are consistent with these assumptions.

5. Results

5.1. Main hypotheses

In this section, we show that changes in local retail options have a substantial effect on the types of products bought online. Table 2 presents our main results.

We first examine how offline store entry influences online buyer behavior through changes in *convenience*. Hypothesis 1 implies that, because of improved convenience, entry by discount stores and large bookstores should decrease purchases of popular books that are more easily available offline. This is exactly what we find. The first 10 rows of column (1) shows how new store entry decreases the likelihood

of purchasing relatively popular books online. Rows 1 and 2 show that discount store and large bookstore entry decrease the likelihood of a local top 10 appearance by books in the national top 150 by 3.2 and 3.4 percentage points respectively; these results are significant at the 1% level. This is a substantial effect, given that the average likelihood of a book appearing in a local top 10 is 3.5%. The negative correlation between new store entry and the local ranking of nationally very popular books is perhaps the strongest and most robust finding in this paper.

Perhaps surprisingly, Table 2 provides relatively little evidence of a *selection* effect. Hypothesis 2 argues that the marginal effect of store entry over the range of less popular books will be greater for big stores than for discount stores. Our test of the selection effect relies on the examination of the difference between discount store and large bookstore entry. In particular, we argue that selection would imply the entry interaction coefficient on books with national popularity of 5001-15,000 should be more negative for large bookstores than for discount stores. We focus on these books because they are likely to be stocked in large bookstores but not in discount stores. In Table 2, we do not find evidence consistent with the selection hypothesis.

We next examine the effects of a change in *price* on product choice. As expected, an increase in Amazon price relative to the list price decreases the likelihood that a book will be purchased online, relative to other books (Hypothesis 3a). The coefficient on relative price in column (1) is negative and statistically significant at the 1% level. Its magnitude of -0.0237 implies that a one standard deviation increase in relative price—equivalent to a 14.34 percentage point deviation from offline price—decreases the likelihood of a book appearing in a local top 10 by 0.3 percentage points. This change is substantial when compared to an average probability of a book appearing in a local top 10 of 3.5%. Thus, as predicted by Hypothesis 3a, discounts increase relative sales.

However, consumers become less sensitive to discounts with the entry of discount stores and large bookstores (Hypothesis 3b). Column (1) of Table 2 shows that the coefficients on the interaction of relative price with discount stores and large bookstores are 0.0147 and 0.0183 respectively; both are statistically significant at the 1% level. Thus while the combined effect (direct plus interaction) implied

by these coefficient estimates of a change in Amazon relative price is -0.0237 in the absence of retailer entry, it is -0.0090 and -0.0054 when discount stores and large bookstores enter, respectively. In other words, in the presence of discount store or large bookstore entry, a one standard deviation increase in price is associated with a 0.1 percentage point decline in the likelihood of a book appearing in the local top 10. Thus, we have significant evidence in support of Hypothesis 3b: as offline discounts increase due to the entry of new retailers, consumers are less sensitive to online discounts. Still, the magnitude of the effect of entry on prices is lower than the effect on convenience as measured by a one standard deviation change.

The R-squared of our baseline model is 0.06. There are two reasons for these relatively low measures. First, these R-squared values are for the "within" (differenced) fixed effect estimator. If we had estimated the fixed effects instead of differencing them out, the measured R-squared would be much higher. However, this latter model is both inconsistent (due to the small number of time periods) and computationally intractable (due to the large number of fixed effects to estimate). Second, our dependent variable is binary, rather than continuous. It is not uncommon to have low R-squared for limited dependent variable models such as this one (e.g. Athey and Stern 2002).

Table 2 demonstrates how entry by offline retailers influences online buyer behavior due to changes in offline price, convenience, and selection. Improvements in convenience cause consumers to shift from buying popular products to less popular products and unpopular products online. However, new store entry also causes online buyers to be less sensitive to changes in online price. Since Amazon discounts best-selling products most heavily, this means that new store entry causes consumers to shift away from popular products due to both convenience and price effects.

Figure 2 shows the net effect of convenience and selection on buyer behavior. The graph shows the change in the likelihood of a local top 10 appearance due to entry by discount or large bookstores. For example, the marginal effect represented by the discount store bar for products in the national top 150 was calculated by computing the change in the likelihood of appearance for books in this range with and without a discount store within 5.4 miles. Discount store and large bookstore entry cause consumers to

shift away from buying the most popular books (those whose national sales ranks are below 150) and shift toward purchasing less popular books.

5.2. Robustness Checks

In columns (2) through (4) of Table 2 we present three robustness checks. In column (2) we show the results of changing our entry radius from 5.4 to 20 miles. This threshold was applied based on the findings of Brynjolfsson and Smith (2000) who find that 8% of consumers live more than 20 miles away from the nearest general selection bookstore. The results are qualitatively similar to those in column (1). In column (3) we examine the use of a continuous measure of book popularity, log (Sales Rank), rather than a spline. While this measure does not allow us to examine Hypothesis 2 on selection, the convenience and price results are qualitatively unchanged.

In column (4) we explore a different choice set definition. When deciding on the size of the set of books to include in the choice set, we had to make a tradeoff between two competing objectives. On the one hand, to identify selection, we wanted to make the size of the choice set as large as possible. On the other hand, if we made the choice set too large, then we would have many products that are rarely in a local top 10. Since our product-location fixed effects rely on differencing dependent and independent variables from mean values, this too is unappealing. As noted above, we finally settled on a choice set of 300. To examine the robustness of our results to a larger choice set, we re-estimated our model using a choice set of 1000 products. The core results are qualitatively similar to those in column (1) although we lose some statistical significance. We believe this is because the fixed effects are inappropriate for those products for which we have a very small number of observed appearances in a local top 10.

5.3. Results for comparative statics hypotheses

In this section we examine how different market and product characteristics influence our findings related to price, convenience, and selection. Table 3 presents our results

In Hypothesis 4a, we argued that store entry should decrease purchases of popular books by more in small markets than in large markets. This is exactly what we find for entry by discount stores in

columns (1) and (2) of Table 3. Entry by discount stores decreases the likelihood of books in the top 150 by 4.1 percentage points in small locations, compared to a decline of 3.0 percentage points in large markets. However, entry by large bookstores has little effect on consumer behavior in small markets, even for the most popular books. This is not surprising, as there is little entry by large bookstores in these locations (Table 1). Consequently, the test has little power.

Hypothesis 4b argues that large markets and markets with heterogeneous tastes are more likely to show a selection effect. Columns (1) through (4) of Table 3 show evidence supporting this. Columns (1) and (2) show that large bookstore entry decreases the likelihood of a local top 10 appearance by books in the top 5000-15,000 by 0.5 percentage points (5% significance) in large locations, while discount stores have no discernible impact on purchase decisions. These coefficient estimates are significantly different from one another at the 5% level. In contrast, in small locations neither discount store entry nor large bookstore entry has any discernible impact on the purchases of lower ranked books.

Columns (3) and (4) present our results for university and non-university towns. Our results on the interaction coefficients of entry with the national top 150 dummy show that new store entry causes both university and non-university towns to substitute away from the most popular books. However, new store entry in university towns also causes consumers to substitute away from less popular books: Large bookstore entry decreases the likelihood of a local top 10 appearance by books in the top 5000-15,000 by 0.4 percentage points (5% significance). This is not true in towns without a university. The university town results also fulfill the second criteria for our selection hypothesis (Hypothesis 2): the coefficient on large bookstores is more negative than that for discount stores for less popular books. For the interaction between the national top 5001-15,000 books, these coefficients are significantly different from one another at the 5% level. No such differences are evident in non-university towns. In short, it appears that consumers in university towns and in large cities (where consumers may prefer relatively less popular books) use online retailers to achieve better selection, while consumers in small towns and towns without a university do not.

In Hypothesis 5, we argued that our results on convenience, selection, and price would be

stronger in locations in which sales taxes are levied. The results in columns (5) and (6) of Table 3 strongly support this assertion. In each of rows 1 through 10, the interaction of our popularity dummies with our entry variables is larger (in absolute value) for locations with online sales taxes than for states without such sales taxes. Moreover, in three cases, these interactions were insignificant for locations without sales taxes and significant for locations with sales taxes. As an example, entry by large bookstores decreases the likelihood that books in the national top 150 would appear in the local top 10 by 10.0 percentage points in sales tax locations (1% significance), while only decreasing the likelihood of local top 10 appearance by 3.2 percentage points in other locations (1% significance). The effect of convenience is clearly stronger in sales tax locations than in locations without sales taxes. Columns (5) and (6) also provide evidence that the effects of selection are stronger in sales tax locations. Large bookstore entry decreases the likelihood of a local top 10 appearance by books in the national top 5000-15,000 by 3.7 percentage points in sales tax locations, while discount store entry has no discernible impact on purchase decision in this category of books. In contrast, large bookstore entry has no discernible impact on local top 10 appearance for this category of books in other locations. Furthermore, for two of our three price coefficients, the absolute value for sales tax locations is also larger than for locations without sales taxes. In sum, we find evidence that the effects of convenience and selection are significantly stronger in sales tax locations, and limited evidence that the price effects are stronger in such locations.

Finally, Hypothesis 6 argued that convenience, selection, and price effects should be stronger for fiction books than for non-fiction books. Columns (7) and (8) of Table 3 support this hypothesis. They suggest that our results on convenience are largely driven by fiction books. The coefficient estimates on the interaction between the top 150 and discount store and large bookstore entry are -0.1110 and -0.1303 for fiction books (1% significance) and only -0.0035 (1% significance) and -0.0004 (not significant) for non-fiction books. The magnitudes of all of the other interactions on the popularity dummies for fiction books are larger (in absolute value) than those for non-fiction books. Similarly, the effects of price are stronger for fiction books than for non-fiction books. For selection, although for fiction books we do find that our point estimate of the interaction of unpopular books with large bookstore entry is larger than that

of the interaction with discount store entry, the coefficient estimate for large bookstore entry is statistically insignificant. In sum, we find support that the convenience and price effects are stronger for fiction books than for nonfiction books.

6. Discussion and Managerial Implications

6.1 Implications for research

Geography matters for electronic commerce. In fact, it matters a lot. Prior research in electronic commerce has emphasized how electronic markets can improve consumer welfare through, for example, lower search costs (e.g., Brynjolfsson and Smith 2000), increased product variety (e.g., Brynjolfsson, Hu, and Smith 2003, Ghose, Smith and Telang 2006), and better product information through online world-of-mouth (e.g., Dellarocas 2003). At the risk of oversimplifying, the virtual world improves consumer welfare by easing constraints that are imposed by the physical world. However, the relative importance of each of these constraints varies across locations in the physical world. While a rich theoretical and empirical literature examines the impact of the online channel on offline choices (e.g., Jeffers and Nault 2005; Brynjolfsson and Smith 2000), in our view the importance of the physical world to consumer behavior in the virtual world has not been widely documented. This paper represents one step in developing an empirical understanding of the role of local offline retailers on online behavior. Table 4 summarizes our main results.

We also provide empirical support for the assumptions of a widely used theoretical modeling framework, the circular city model adjusted to include a direct channel (e.g. Balasubramanian 1998). We find that variables and parameters in these models such as offline transportation cost, online product fit, and the prices of online and offline retailers interact to determine consumers channel choice in a way that is consistent with these models. However, our results also identified a potentially valuable extension to these models. Our results suggest that in certain environments retailer decisions of whether to stock less popular products had a significant impact on the manner in which consumers substituted between the online and offline channels.

Our emphasis on the roles of convenience and selection for online buyer behavior also advances a small empirical literature on online-offline channel substitution that has thus far focused primarily on the ability of the online channel to offer lower prices (e.g., Prince 2006). In fact, our results showed that (travel cost) convenience was the single biggest factor influencing online-offline substitution. More broadly, though prior work has shown that convenience may influence the dollar value of spending on internet commerce (Sinai and Waldfogel 2004), to our knowledge no prior work has examined how convenience shapes consumers' product choices online. Our results suggest that future work on channel substitution should emphasize the importance of convenience to online buying behavior.

We find the impact of selection is relatively small, and is isolated in locations such as large markets or university towns where tastes may be more heterogeneous. To be clear, this is not to say that our results imply that the impact of the Internet on improving consumer welfare through increased selection is small. Moreover, it is not our goal to identify consumer shifts to the “long tail” (Anderson 2006; Elberse and Oberholzer-Gee 2006) of rare products traditionally not supplied in the offline world. Rather, we seek to observe channel substitution among products that are supplied by offline retailers. Our results show there is relatively less evidence of consumers substituting between lower ranked books available in large bookstores and the same lower ranked books available in the online channel. We speculate that this may be because these two channels provide different kinds of “convenience” to consumers. While offline bookstores have higher travel costs, they do make it easier for consumers to browse through various categories of books. Our results on fiction books support this possibility, suggesting that browsers who use large offline stores may represent a slightly different market than online users, leading to limited online-offline substitution for less popular items.

Finally, we make a small methodological contribution. We observe a sales measure of some products (online books), and the presence of rival products (offline stores), but not the sales of the rival products. Thus, at a broader level, our paper provides a new modeling framework for examining how the presence of one product might influence sales of another.

6.2 Implications for managers

For both online and offline firms, our results provide important information on how consumers substitute across online and offline channels, a topic on which there is still relatively little empirical analysis. Entry by new offline firms will shift consumers' online behavior, and so electronic retailers should adjust accordingly. Online firms should view geographic expansion by discount stores and by large bookstores as a threat. Further, our results suggest that the increasing diffusion of discount stores and large bookstores across the US may lead to consumers using the Internet channel mainly for less popular products. This is in addition to recent evidence that consumers are increasingly using Internet channels to purchase relatively unpopular products, or those products on the "long tail" of the sales distribution (e.g., Anderson 2006). Our results also suggest that advertisements and informational displays by online retailers should be targeted to consumers based upon local retail supply conditions. In this context, we are already seeing the advent of location based geo-targeting by online firms in both traditional advertisements as well as search-engine based sponsored advertisements. For offline entrants, our results show that competition will depend not only upon the number of local stores, but also upon consumer adoption of electronic channels.

6.3 Limitations of Research

Our research represents the first attempt to integrate the effects of convenience, selection, and price in an empirical model of online buyer behavior. However, as with any empirical work, the data that we bring to bear has some limitations. For one, we are restricted in that we only observe the top ten products in each location. Thus, though there is considerable heterogeneity in top products across locations and many observed purchases of unpopular goods, we are limited in our ability to make inferences about purchases of very unpopular products.

Also, we are unable to observe the decision to use the online channel. Our inferences are based on changes in the popularity of the products consumers purchase across locations and over time, but we have no information on the binary decision of a consumer to use the online channel at all.

Our results focus on entry of offline retailers to measure how cross-sectional variation in local supply conditions influences online consumer behavior. We pursue this approach because any cross-sectional study that examines the relationship between consumer behavior and local supply will be confounded by the possibility that tastes vary across locations. Therefore, our conclusions assume that time-series changes in behavior following entry are similar to differences in behavior in response to cross-sectional local retail variation. In more formal language, we use within-location variation to make inferences about between-location variation.

Our data do not allow us to observe the exact mechanism that drives the substitution; we simply observe that offline competition changes the distribution of products bought online. We observe that where you are offline affects what you buy online, but not exactly why. Though our empirical findings may be consistent with several alternative mechanisms, this does not influence the primary findings of this paper: that consumers substitute between online and offline channels to obtain better convenience, selection, and price.

Last, we examine online behavior for only one particular product: books. Due to the large market share that Amazon has in books and the source of our data, we argue that this is a good place to begin to understand how differences in supply conditions shape variance in online buyer behavior. However, it may be interesting to explore how the distribution of sales is affected by offline supply conditions for other products. In particular, this may be important for products with higher prices which typically have higher involvement due to greater perceived risks from product uncertainty or with more non-digital attributes that consumers may prefer to experience before purchasing online (Lal and Sarvary 1999).

6.4. Conclusion

Utilizing a unique panel data set of online purchases of books by consumers across urban and rural locations in the US, we examine how geography 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

those with more options.

We draw on the existing theoretical literature to motivate how variance in convenience, selection, and price across local markets may influence consumer behavior online. Our empirical results support the hypotheses motivated by this literature. In particular, we demonstrate that offline convenience and selection play a significant role in driving consumer decisions to use the online or offline channel. These effects are further reinforced by market and product characteristics such as market size and diversity, online taxes and product type.

More generally, we provide evidence that the online channel is used for different reasons across locations. Despite the fact that each consumer (irrespective of location) receives the same product and service when purchasing online, the reasons for buying the product online differ by location. Future research can extend this stream of work by looking at disaggregated data purchases at the individual consumer level to identify other reasons why consumer purchases in electronic markets vary across locations.

References

- Ai, C. and E. Norton. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters.* 80, 123-129.
- Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More.* Hyperion Press, New York, NY.
- Athey, S., and S. Stern. 2002. The Impact of Information Technology on Emergency Health care Outcomes. *RAND Journal of Economics* 33(3), 399–432.
- Bajari, P., J.T. Fox, and S. Ryan. 2006. Evaluating Wireless Carrier Consolidation Using Semiparametric Demand Estimation. NBER Working Paper # 12425.
- Bakos, J.Y. 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12), 1676-1692.
- Balasubramanian, S. 1998. Mail versus mall: a strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science* 17(3), 181-195.
- Balasubramanian, S., P. Konana and N. Menon. 2003. Customer Satisfaction in Virtual Environments: A Study of Online Investing. *Management Science*. 49(7), 871-889.

- Baye, Michael, John Morgan, and Patrick Scholten. 2007. Information, Search, and Price Dispersion," in *Handbook of Economics and Information Systems* (T. Hendershott, ed.), Elsevier Press, forthcoming.

- Bresnahan, T. and P. Reiss. 1991. Entry and Competition in Concentrated Markets. *Journal of Political Economy* 99, 977-1009.

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

- Brynjolfsson, E. and M. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46(4), 563-585.

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

- Cairncross, F., 1997. The Death of Distance. Cambridge, MA: Harvard University Press.

- Cheng, J. and B. Nault. 2006. Internet Channel Entry: Retail Coverage and Entry Cost Advantage. Forthcoming, *Information Technology Management.*

- Chiou, L. 2005. Empirical Analysis of Retail Competition: Spatial Differentiation at Wal-Mart, Amazon.com, and their Competitors. Working Paper, Occidental College.

- Clemons, E., Hann, I., and L. Hitt. 2002. Price Dispersion and Differentiation in Online Travel: An Empirical Investigation, *Management Science* 48(4), 534-550.

- Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science* 49(10), 1407-1424.

- Dewan, R., M. Freimer, and A. Seidmann. 2000. Organizing Distribution Channels for Information Goods on the Internet. *Management Science* 46(4): 483-495.

- Druehl, C. and E.L. Porteus. 2005. Online versus Offline Price Competition: Market Outcomes and the Effect of Internet Shopping Penetration. Working Paper, Smith School of Business, University of Maryland.

- Elberse, A., and F. Oberholzer-Gee. 2006. Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales. Harvard Business School Working Paper, No. 07-015.

- Ellison, G. And S. Ellison. 2006. Internet Retail Demand: Taxes, Geography, and Online-Offline Competition, NBER Working Paper 12242.

- Forman, C., A. Goldfarb, and S. Greenstein. 2005. How did location affect adoption of the commercial Internet: global village vs. urban leadership. *Journal of Urban Economics* 58(3), 389-420.

- Forman. C, A. Ghose and B. Wiesenfeld. 2006. A Multi-Level Examination of the Impact of Social Identities on Economic Transactions in Electronic Markets. NYU CeDER Working Paper # 06-09.

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

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

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

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

- Jeffers, P. and B.R. Nault. 2005. Why Online Retail Competition Does Not Always Benefit Consumers. Working Paper, Haskayne School of Business, University of Calgary.

- Lal, R. and M. Sarvary 1999. When and How is the Internet Likely to Decrease Price Competition. *Marketing Science*, 18(4), 485-503.

- Prince, J. 2006. The Beginning of Online/Retail Competition and its Origins: An Application to Personal Computers. Forthcoming, *International Journal of Industrial Organization.*

- Riggins, F.J. 2004. A Multichannel Model of Separating Equilibrium in the Face of the Digital Divide. *Journal of Management Information Systems* 21(2), 161-179.

- Salop, S.C. 1979. Monopolistic competition with outside goods. *Bell Journal of Economics* 10, 141-156.

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

- Viswanathan, S. 2005. Competing Across Technology-Differentiated Channels: The Impact of Network Externalities and Switching Costs. *Management Science* 51(3), 483-496.

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

- Wolfinbarger, M., M.C. Gilly. 2001. Shopping Online for Freedom, Control, and Fun. *California Management Review* 43(2), 34-55.

Table 1: Summary Statistics for Books

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
BY LOCATION-PRODUCT-MONTH					
Dummy for Top 10 in location	4,062,326	0.0347	0.1831	0	1
Relative Price	4,062,326	-0.2654	0.1434	-0.6	0
Top 150 Books	4,062,326	0.1711	0.3766	0	1
Top 151-500 Books	4,062,326	0.1737	0.3789	0	1
Top 501-1500 Books	4,062,326	0.1538	0.3608	0	1
Top 1501-5000 Books	4,062,326	0.1351	0.3418	0	1
Top 5001-15,000 Books	4,062,326	0.1296	0.3358	0	1
Dummy for missing price information	4,062,326	0.0644	0.2454	0	1
Average rating	4,062,326	4.1098	0.5617	1.5	5
Log(days since launch)	4,062,326	6.5007	1.4946	0	9.8268
Dummy for missing elapsed date information	4,062,326	0.0259	0.1588	0	1
Log(Number of reviews)	4,062,326	4.9545	1.4596	0.6931	8.6500
Discount Store Entry within 5.4 miles	4,062,326	0.0807	0.2724	0	1
Large Bookstore Entry within 5.4 miles	4,062,326	0.0165	0.1274	0	1
BY LOCATION					
Discount store openings in all locations	1501	0.1639	0.3703	0	1
Discount store openings in small locations	143	0.0979	0.2982	0	1
Discount store openings in large locations	416	0.2067	0.4054	0	1
Large bookstore openings in all locations	1501	0.0466	0.2109	0	1
Large bookstore openings in small locations	143	0.0210	0.1438	0	1
Large bookstore openings in large locations	419	0.0745	0.2629	0	1
Location has a university	1501	0.4444	0.4971	0	1

Note: Unit of observation in top half of table is a location-product-month. Unit of observation in the bottom half of the table is a location.

Table 2: Main Book Results—Difference in Difference on store entry

	(1)	(2)	(3)	(4)
	5.4 Miles	20 Miles	Sales Rank	Large Choice Set
(Top 150 Books)*	-0.0320	-0.0372	. . .	-0.0343
(Discount Store Entry within 5.4 miles)	(0.0012)**	(0.0008)**	. . .	(0.0011)**
(Top 150 Books)*	-0.0339	-0.0343	. . .	-0.0387
(Large Bookstore Entry within 5.4 miles)	(0.0025)**	(0.0011)**	. . .	(0.0023)**
(Top 151-500 Books)*	-0.0034	-0.0061	. . .	-0.0018
(Discount Store Entry within 5.4 miles)	(0.0008)**	(0.0005)**	. . .	(0.0005)**
(Top 151-500 Books)*	-0.0029	-0.0045	. . .	-0.0017
(Large Bookstore Entry within 5.4 miles)	(0.0020)	(0.0008)**	. . .	(0.0012)
(Top 501-1500 Books)*	-0.0060	-0.0080	. . .	0.0018
(Discount Store Entry within 5.4 miles)	(0.0006)**	(0.0003)**	. . .	(0.0003)**
(Top 501-1500 Books)*	-0.0022	-0.0048	. . .	0.0041
(Large Bookstore Entry within 5.4 miles)	(0.0016)	(0.0006)**	. . .	(0.0007)**
(Top 1501-5000 Books)*	-0.0082	-0.0084	. . .	-0.0007
(Discount Store Entry within 5.4 miles)	(0.0009)**	(0.0004)**	. . .	(0.0003)+
(Top 1501-5000 Books)*	-0.0074	-0.0067	. . .	0.0005
(Large Bookstore Entry within 5.4 miles)	(0.0025)**	(0.0009)**	. . .	(0.0009)
(Top 5001-15,000 Books)*	-0.0019	-0.0020	. . .	0.0006
(Discount Store Entry within 5.4 miles)	(0.0007)**	(0.0003)**	. . .	(0.0002)*
(Top 5001-15,000 Books)*	-0.0022	-0.0021	. . .	0.0009
(Large Bookstore Entry within 5.4 miles)	(0.0018)	(0.0006)**	. . .	(0.0005)+
Log(Sales Rank)*	. . .	. . .	0.0034	. . .
(Discount Store Entry within 5.4 miles)	. . .	. . .	(0.0002)**	. . .
Log(Sales Rank)*	. . .	. . .	0.0035	. . .
(Large Bookstore Entry within 5.4 miles)	. . .	. . .	(0.0004)**	. . .
Relative price	-0.0237	-0.0268	-0.0138	-0.0149
	(0.0007)**	(0.0008)**	(0.0023)**	(0.0004)**
(Relative Price)*	0.0147	0.0107	0.0206	0.0166
(Discount Store Entry within 5.4 miles)	(0.0022)**	(0.0010)**	(0.0106)*	(0.0012)**
(Relative Price)*	0.0183	0.0145	0.0174	0.0207
(Large Bookstore Entry within 5.4 miles)	(0.0061)**	(0.0021)**	(0.0228)	(0.0031)**
Top 150 Books	0.006	0.0180	. . .	0.0141
	(0.0004)**	(0.0005)**	. . .	(0.0003)**
Top 151-500 Books	-0.0124	-0.0108	. . .	-0.0075
	(0.0003)**	(0.0003)**	. . .	(0.0002)**
Top 501-1500 Books	-0.0061	-0.0042	. . .	-0.0011
	(0.0003)**	(0.0003)**	. . .	(0.0001)**
Top 1501-5000 Books	-0.0066	-0.0045	. . .	-0.0007
	(0.0002)**	(0.0002)**	. . .	(0.0001)**
Top 5001-15,000 Books	-0.0041	-0.0033	. . .	-0.0002
	(0.0001)**	(0.0001)**	. . .	(0.0001)**
Observations	4,062,326	4,062,326	4,062,326	9,446,406
Number of FEs	981,255	981,255	981,255	2,941,782
R-squared	0.0580	0.0582	0.0602	0.0355
Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Log(Sales Rank) (Column 3 only)	-Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry within 5.4 (or 20) miles - Large Bookstore Entry within 5.4 (or 20) miles -Product-location fixed effects (differenced out)			

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

Table 3: Results split by location size, whether the location has a university, whether the location has a sales tax, and whether the book is fiction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Locations Under 100k	Locations over 1 million	Locations with a university	Locations without a university	No Sales Tax Online	Sales Tax Online	Fiction books only	Books that are not fiction
(Top 150 Books)*	-0.0409	-0.0299	-0.0348	-0.0291	-0.0315	-0.0458	-0.1110	-0.0035
(Discount Store Entry within 5.4 miles)	(0.0045)**	(0.0023)**	(0.0016)**	(0.0018)**	(0.0012)**	(0.0061)**	(0.0045)**	(0.0011)**
(Top 150 Books)*	-0.0067	-0.0389	-0.0426	-0.0276	-0.0320	-0.1004	-0.1303	-0.0004
(Large Bookstore Entry within 5.4 miles)	(0.0176)	(0.0045)**	(0.0039)**	(0.0033)**	(0.0025)**	(0.0253)**	(0.0108)**	(0.0023)
(Top 151-500 Books)*	-0.0071	0.0014	-0.005	-0.0012	-0.0033	-0.0073	-0.0315	-0.0016
(Discount Store Entry within 5.4 miles)	(0.0030)*	(0.0016)	(0.0010)**	(0.0013)	(0.0009)**	(0.0032)**	(0.0036)**	(0.0009)+
(Top 151-500 Books)*	-0.0025	-0.0073	-0.0118	0.0034	-0.0022	-0.0466	-0.0368	0.0009
(Large Bookstore Entry within 5.4 miles)	(0.0142)	(0.0034)*	(0.0027)**	(0.0027)	(0.0019)	(0.0238)*	(0.0086)**	(0.0020)
(Top 501-1500 Books)*	-0.0065	-0.0024	-0.0075	-0.0040	-0.0057	-0.0123	-0.0299	-0.0058
(Discount Store Entry within 5.4 miles)	(0.0028)*	(0.0013)+	(0.0008)**	(0.0010)**	(0.0006)**	(0.0022)**	(0.0025)**	(0.0007)**
(Top 501-1500 Books)*	0.0112	-0.0061	-0.0097	0.0033	-0.0016	-0.0472	-0.0242	-0.0012
(Large Bookstore Entry within 5.4 miles)	(0.0115)	(0.0026)*	(0.0021)**	(0.0024)	(0.0016)	(0.0214)*	(0.0068)**	(0.0019)
(Top 1501-5000 Books)*	-0.0072	-0.0088	-0.0087	-0.0074	-0.0079	-0.0133	-0.0220	-0.0126
(Discount Store Entry within 5.4 miles)	(0.0050)	(0.0019)**	(0.0011)**	(0.0015)**	(0.0009)**	(0.0036)**	(0.0019)**	(0.0012)**
(Top 1501-5000 Books)*	0.0172	-0.0105	-0.0141	-0.0022	-0.0069	-0.0564	-0.0180	-0.0124
(Large Bookstore Entry within 5.4 miles)	(0.0116)	(0.0040)**	(0.0038)**	(0.0034)	(0.0025)**	(0.0243)*	(0.0052)**	(0.0031)**
(Top 5001-15,000 Books)*	-0.0006	0.0014	-0.0023	-0.0013	-0.0018	-0.0031	-0.0029	-0.0007
(Discount Store Entry within 5.4 miles)	(0.0037)	(0.0016)	(0.0009)**	(0.0011)	(0.0007)**	(0.0031)	(0.0014)*	(0.0007)
(Top 5001-15,000 Books)*	0.0167	-0.0049	-0.0043	-0.0008	-0.0021	-0.0368	-0.0034	-0.0011
(Large Bookstore Entry within 5.4 miles)	(0.0157)	(0.0024)*	(0.0019)*	(0.0031)	(0.0018)	(0.0042)**	(0.0035)	(0.0019)
Relative price	-0.0221	-0.0275	-0.0258	-0.0220	-0.0234	-0.0319	-0.0026	-0.0194
	(0.0023)**	(0.0015)**	(0.0012)**	(0.0010)**	(0.0008)**	(0.0040)**	(0.0015)+	(0.0007)**
(Relative Price)*	0.0215	0.0241	0.0159	0.0136	0.0152	0.0068	0.0676	0.0245
(Discount Store Entry within 5.4 miles)	(0.0102)*	(0.0046)**	(0.0028)**	(0.0035)**	(0.0022)**	(0.0080)	(0.0056)**	(0.0023)**
(Relative Price)*	0.0160	0.0041	0.0130	0.0226	0.0157	0.0696	0.0727	0.0317
(Large Bookstore Entry within 5.4 miles)	(0.0196)	(0.0090)	(0.0078)+	(0.0089)*	(0.0061)**	(0.0267)**	(0.0157)**	(0.0060)**
Observations	386,551	1,119,715	1,811,105	2,251,221	3,904,989	157,337	1,207,420	2,854,906
Number of FEs	93,393	271,419	437,108	544,147	943,497	37,758	306,160	675,095
R-squared	0.0535	0.0588	0.0606	0.0559	0.0581	0.0558	0.0838	0.0440

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects and all the same variables as in Table 2.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 4: Main Hypotheses and Summary of Results

Hypothesis	*Variable*	*Prediction*	*Intuition*	*Supported?*	*Location*
1 (Convenience)	(Very popular books)* (Large bookstore/ Discount store entry)	Negative	With an increase in the number of stores, more popular products are bought offline	Supported	Table 2
2 (Selection)	(Less popular books)* (Large bookstore entry)	Negative, more so than the interaction with discounter entry	With an increase in the number of large bookstores, more unpopular products are bought offline.	Not Supported in full data set	Table 2
3a (Price)	Relative price	Negative	Books that are discounted will be bought more frequently than books that are not discounted.	Supported	Table 2
3b (Price)	(Relative price)* (Large bookstore/ Discount store entry)	Positive	Items that are discounted online are typically discounted offline. Therefore, the impact of online discounts will be tempered by local retail stores.	Supported	Table 2
4a (Market size and convenience)	(Very popular books)* (Large bookstore/ Discount store entry)	Negative, more so in small markets	Convenience effects are stronger in small markets where distance to stores is larger.	Supported	Table 3, columns 1 and 2
4b (Market size and selection)	(Less popular books)* (Large bookstore entry)	Negative, more so in large markets and university towns	Selection effects are stronger in large markets/university towns where consumer preferences are more diverse.	Supported	Table 3, columns 1 through 4
5 (Taxes and convenience)	(Very popular books)* (Large bookstore/ Discount store entry)	Negative, more so in locations with online sales tax	Convenience effects are stronger when fit of the online channel is worse due to sales taxes.	Supported	Table 3, columns 5 and 6
5 (Taxes and selection)	(Less popular books)* (Large bookstore entry)	Negative, more so in locations with online sales tax	Selection effects are stronger when fit of the online channel is worse due to sales taxes.	Supported	Table 3, columns 5 and 6
5 (Taxes and price)	(Relative price)* (Large bookstore/ Discount store entry)	Positive, more so in locations with online sales tax	Price effects are stronger when fit of the online channel is worse due to sales taxes.	Partially supported	Table 3, columns 5 and 6
6 (Fiction books and convenience)	(Very popular books)* (Large bookstore/ Discount store entry)	Negative, more so for fiction books	Convenience effects are stronger when fit of the online channel is worse for fiction books.	Supported	Table 3, columns 7 and 8
6 (Fiction books and selection)	(Less popular books)* (Large bookstore entry)	Negative, more so for fiction books	Selection effects are stronger when fit of the online channel is worse for fiction books.	Partially Supported	Table 3, columns 7 and 8
6 (Fiction books and price)	(Relative price)* (Large bookstore/ Discount store entry)	Positive, more so for fiction books	Price effects are stronger when fit of the online channel is worse due to sales taxes.	Supported	Table 3, columns 7 and 8

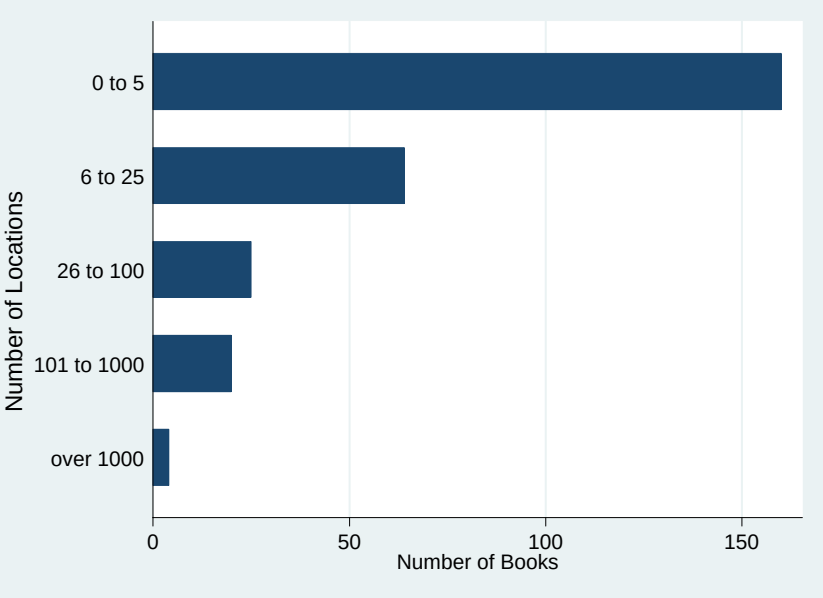

Figure 1: Number of Locations that a Book is in the Top 10--May 2005

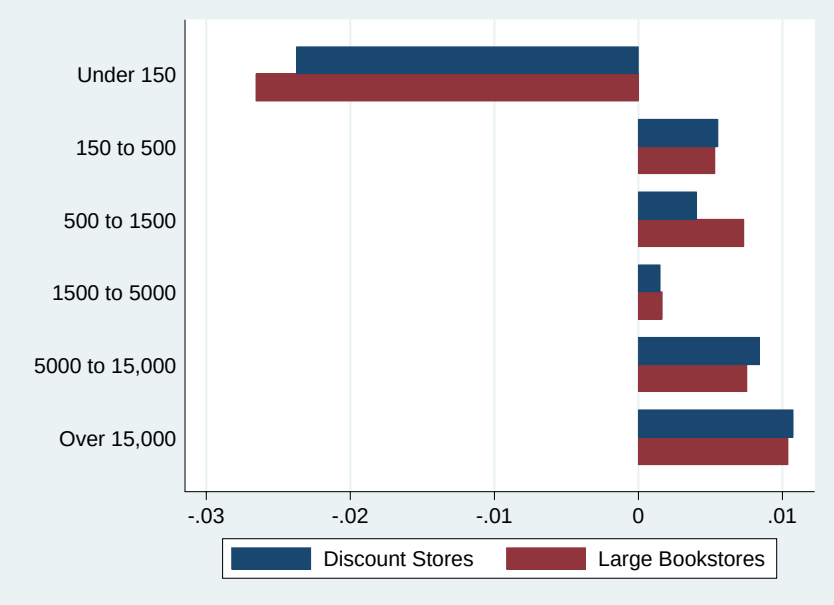

Figure 2: Discount Store/Specialty Store Comparison
Marginal Effect of Store Entry by Sales Rank,
(Based on Table 2 column (1))