NET Institute*

www.NETinst.org

Working Paper #06-15

November 2006

Geography and Electronic Commerce: Measuring Convenience, Selection, and Price

Chris Forman,Tepper School of Business
Anindya Ghose, Stern School of Business, New York University
Avi Goldfarb, Rotman School of Management, University of Toronto

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Geography and Electronic Commerce: Measuring convenience, selection, and price[1]

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

November 2006

NET Institute Working Paper

[1] Financial support from the NET Institute www.NETinst.org is gratefully acknowledged.

Geography and Electronic Commerce: Measuring convenience, selection, and price

Abstract

We develop a formal model of online-offline retail channel substitution to identify three factors that drive consumers to purchase online: convenience, selection, and price. This model builds hypotheses on how features of offline retail supply impact online purchasing. We then 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.

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 the 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.[2]

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. An 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 an analytical framework to provide insight into the nature of online-offline channel substitution. In this framework, a consumer's choice of channel varies based upon differences in offline selection, convenience, and price. The model predicts that consumers who are located closer to retail stores are more likely to buy popular products offline relative to those who are further from retail stores. Similarly, those located in markets with broader selection are more likely to buy relatively unpopular products offline. This framework provides three testable hypotheses that each relate to one of convenience, selection, and price.

[2] 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.

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. These results are particularly strong for fiction books. We also find evidence that consumers substitute to online channels to obtain lower *prices*, especially in cities with populations over 1 million people. Overall, we do not find any 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, in larger cities, and in states that levy taxes on online sales. When a Barnes & Noble opens in these locations, consumers buy fewer books that are nationally ranked between 5000 and 15,000 in the online channel. Wal-Mart stores, which do not typically carry these books, do not have this effect. More generally, we find that geography matters: local retail supply characteristics are important determinants of online behavior.

1.1 Related Literature and Contribution

This paper advances several streams of research. First, it contributes to recent research that examines consumer substitution between online and offline channels (Chiou 2005; Goolsbee 2000; Ellison and Ellison 2006, Prince 2006).[3] Of these the two most relevant papers are Chiou (2005) and Ellison and Ellison (2006). Chiou (2005) finds that conditional on price and distance, the average consumer still prefers Wal-Mart over most other stores. Ellison and Ellison (2006) show that tax avoidance may be an important contributor to e-retail activity and find that geography matters in two ways. First, there is some evidence that consumers prefer purchasing from firms in nearby states to benefit from faster shipping times. Second, there is evidence of a separate preference for buying from in-state firms. Recent work has also found that online-offline channel substitution has the potential to substantially alter the market structure of offline retail industries (Emre, Hortacsu, and Syverson 2005).

In contrast to this prior work that has focused on cross-price elasticity between online and offline channels, our research focuses on how changes in offline price, convenience, and selection influence consumer choice of products online. 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 (Brown and Goolsbee 2002; Brynjolfsson and Smith 2000; Clemons, Hann, and Hitt 2002; Granados, Gupta and Kauffman 2006, Ghose and Sundararajan 2006), and demonstrated that consumers benefit from lower prices in the online channel. A related line of research has shown that Internet retailing improves consumer welfare primarily by lowering search costs. 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 the role of Internet technology in reducing the costs associated with distance has been recognized elsewhere (Forman, Goldfarb, and Greenstein 2005; Sinai and Waldfogel 2004), 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

[3] Here we discuss related empirical literature on online-offline channel substitution. For examples of related theory work in this area, see Balasubramaniam (1998), Lal and Sarvary (1999) and Cheng and Nault (2006). Other related theoretical work includes Zhu, Dukes and Singh (2005) who analyze the competition between two spatially differentiated multi-product retailers who encounter entry from a dominant discount retailer.

consumers' channel decision. In doing so, we show how consumers benefit from access to electronic markets. 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).[4] 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 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 provide an analytical framework that leads to three testable hypotheses on convenience, selection and price. Section 3 describes the data. The empirical model is described in Section 4. The analysis along with some robustness checks and extensions are provided in Section 5. We conclude in Section 6 with some discussion of managerial and research implications.

2. Theory

In this paper, we examine variance in consumers' choice of books across locations to identify the factors influencing online-offline channel substitution. In this section, we develop a model to better understand how offline retail options can affect online choices.[5]

[4] Also see Hortacsu, Martínez-Jerez, and Douglas (2006), who examine how geographic location influence trade patterns on eBay and MercadoLibre.

[5] Online behavior may vary across locations for both demand and supply reasons. In our theoretical model we focus on supply of local retail stores. We assume constant preferences across locations. In our empirical work, we add product-location fixed effects to control for differences in consumer preferences across locations.

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

$$q_A = 1-d- b_1p_A + b_2p_B \tag{1}$$

and, the demand function for the online firm is given by

$$q_B = d- b_1p_B + b_2p_A \tag{2}$$

where $0 \leq b_2 < b_1$. Here b_1 and b_2 are the own price and cross-price effects. The parameter *d* represents the intercept of the demand curve, which can be construed as the distance to the offline store from a consumer's location. Intuitively, as *d* increases, consumers face increased transportation costs of traveling to the offline store (for example, due to a decrease in the number of offline retailers). This results in an increased likelihood of purchasing the same product 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 $r \in (0,1)$ and $1-r$ be the parameters representing the popularity of the rare and popular products, respectively. Then, as *r* increases, product 1 becomes relatively more popular and product 2 becomes relatively less popular. Let Φ denote the probability of availability (i.e. the expected availability from the consumer's perspective) of the unpopular product in the offline store. Offline stores have limited shelf-space and are more likely to stock the relatively more popular products. Without loss of generality, we normalize the costs of each retailer to be equal to zero.

Let the expected demand faced by the offline firm (firm *A*) for the unpopular product be denoted by D_{A1} and the expected demand faced by the offline firm for the popular product be denoted by D_{A2}. Similarly, let D_{B1} and D_{B2} represent the expected demand curves for the online firm (firm B) for the unpopular and popular products, respectively.

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 *C*. *C* denotes the shelf-space or inventory constraint for the offline retailer. It is used to capture the fact that the retailer faces a tradeoff between selling popular versus rare products. So total sales are equal to *C*. Combined with availability Φ, it signifies the likelihood that a consumer who travels to the offline store will find the rare book being stocked by the retailer. This parameter parsimoniously helps us capture that the probability of finding a rare product will vary based on whether it is a discount store or large bookstore because each of these store types has a different capacity for books.

Hence, we have the following constraints for unpopular and popular products, respectively.

$$\Phi C = r\, D_{A1} \tag{3}$$

$$(1-\Phi)C = (1-r)\, D_{A2} \tag{4}$$

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.[6] Therefore, the profit function of the offline firm is given by

$$r\underbrace{(1-d-b_1p_{A1}+b_2p_{B1})}_{D_{A1}}\,p_{A1} + (1-r)\underbrace{(1-d-b_1p_{A2}+b_2p_{B2})}_{D_{A2}}\,p_{A2} +$$
$$\lambda_1(\Phi C - r D_{A1}) + \lambda_2((1-\Phi)C-(1-r)D_{A2})$$

which upon relevant substitution becomes equal to the following expression

$$\pi_1 = \begin{array}{l} r\underbrace{(1-d-b_1p_{A1}+b_2p_{B1})}_{D_{A1}}\,p_{A1} + (1-r)\underbrace{(1-d-b_1p_{A2}+b_2p_{B2})}_{D_{A2}}\,p_{A2} + \\ \lambda_1(\Phi C - r(1-d-b_1p_{A1}+b_2p_{B1})) + \lambda_2((1-\Phi)C - r(1-d-b_1p_{A2}+b_2p_{B2})) \end{array} \tag{5}$$

Similarly, the profit function of the online firm is given by

$$\pi_2 = r\underbrace{(d-b_1p_{B1}+b_2p_{A1})}_{D_{B1}}\,p_{B1} + (1-r)\underbrace{(d-b_1p_{B2}+b_2p_{A2})}_{D_{B2}}\,p_{B2} \tag{6}$$

Hence, in equilibrium the realized demand curve expressions are given by the following equations:

[6] This assumption is supported by considerable anecdotal evidence. 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 example, 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).

$$M_{A1} = \frac{C\Phi}{r} \tag{7}$$

$$M_{A2} = \frac{C(1-\Phi)}{(1-r)} \tag{8}$$

$$M_{B1} = \left(\frac{b_1(r(d(b_1 - b_2) + b_2) - \Phi C b_2)}{r(2b^2_1 - b_2^2)} \right) \tag{9}$$

$$M_{B2} = \left(\frac{b_1((1-r)d(b_1 - b_2) + b_2(1 - r - (1-\Phi)C))}{(2b^2_1 - b_2^2)(1-r)} \right) \tag{10}$$

We use the above equations and profit functions to derive three propositions that can be interpreted as testable empirical hypotheses. Our first proposition describes how increased convenience influences consumers' behavior in the online channel. Proofs of all Propositions are in the Appendix.

Proposition 1: Online purchasing for convenience: *For all $r > \Phi$, as distance to offline stores, d, decreases, the proportion of popular products being bought online also decreases. That is,*

$$\frac{\partial}{\partial d}\left(\frac{M_{B2}}{M_{B1} + M_{B2}} \right) > 0 .$$

Intuitively, with a reduction in the distance to an offline store, consumers derive a higher utility from buying popular products from offline stores since the transportation costs of purchasing products locally will decrease. The products available in offline retail stores tend to be the more popular products. Thus, a reduction in distance to retailers in the offline world will reduce the probability of purchasing popular products online.[7] Our second proposition describes how broader selection influences online behavior.

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

$$\frac{\partial}{\partial \Phi}\left(\frac{M_{B1}}{M_{B1} + M_{B2}} \right) < 0 .$$

Intuitively, if a book is available in the offline market the consumer is more likely to buy it offline

[7] Note that $r > \Phi$ is not very restrictive. It is likely to be true in most cases since the expected availability of a product is generally going to be less than the extent of its popularity. Further note in Proposition 1, we refer to popular books because these are the books that are most likely to be stocked by offline retailers. Changes in distance to offline stores matters less for unpopular books because there is a much lower probability of rare and obscure books being available in the offline world.

and less likely to buy it online, other things equal. On the other hand, as product assortments in the offline world decrease, consumers are less likely to find rare products offline and hence, more likely to buy these products online.

To derive the next result on the effect of prices on consumers' buying behavior across channels, we slightly extend our base model to incorporate the fact that retailers often offer discounts on popular books. We introduce parameters ω_{A2} and ω_{B2} to denote the fixed discount off the retail prices for popular books offered by retailers in the offline and the online channels, respectively. In the book industry, it is very common for the retailer to offer a 30-40% discount on popular books.[8] This leads to the following result.

Proposition 3a: Online purchasing for price: *As online discounts for popular products increase, consumers are more likely to buy them online, leading to higher sales for popular products through the online channel. That is,* $\frac{\partial M_{B2}}{\partial \omega_{B2}} > 0$.

Proposition 3b: Online purchasing for relative price: *The increase in online sales for popular products due to an increase in online discounts will be increasing in the distance to offline stores. That is,* $\frac{\partial^2 M_{B2}}{\partial d \partial \omega_{B2}} > 0$.

Propositions 3a and 3b are based on the well-known intuition that consumers are sensitive to price differences for the same product across the two channels. Proposition 3a simply implies as the price in a channel decreases, demand in that channel is likely to increase. Proposition 3b is similar to a statement about cross-elasticity in that we examine how changes to the total costs of book purchase (including transportation costs) in one channel influence demand in another. In particular, a decrease in price is likely to increase own quantities relatively more than the decrease in quantity demanded in the other channel because the own price effect is stronger than the cross-price effect. Further, a positive cross-price elasticity also implies that offline and online channels compete with each other. Proposition 3b is anecdotally supported by the fact that books that are discounted at Amazon (popular books) tend to be the same as those discounted by offline retailers. Wal-Mart, in particular, typically provides slightly higher discounts than other retailers because of its every day low pricing (EDLP) policy and larger size[9]. Hence, consumer

[8] Typically this is a standard discount percentage on popular books in the best seller list– independent of the list price or the wholesale price (Lee and Png 2004). Therefore, we consider a fixed exogenous discount off the retail price in our model. This also keeps the analysis simple and tractable. Our results are also robust to scenarios where the discount off the retail price is provided in the form of a rebate by the manufacturer. Since this is less likely in the book industry, we chose to adopt the above setup.

[9] Among conventional bookstores, there is evidence that bestseller discounts systematically increase with the store area (Lee and Png 2004).

sensitivity to discounts for popular books in the online world is increasing in the distance to offline retailers such as Wal-Mart.

It is important to note that there can be an alternative to Proposition 3b which suggests that consumer sensitivity to online discounts actually decreases as their distance to an offline store increases. For example, if the distance to the offline firm increases, there would be a greater proportion of people who don't compare across channels ceteris paribus. To allow for this possibility, we do not impose any restrictions on the empirical estimations and let the data set speak for itself.

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. We used a JAVA spider to extract and parse data from Amazon’s website. Amazon is by far the largest online retailer of books with a 70% market share (Forrester 2002). Amazon’s Purchase Circles are specialized best-seller lists that denote the top-selling books by location throughout the US.[10]

The Purchase Circles are organized in multiple layers: first, by state and then within a state, by location. The number of geographic locations that are listed for any given state is very large. For example, the state of Pennsylvania alone lists 71 locations. For each month 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. While previous studies on electronic commerce have used data from Amazon, our use of the Purchase Circles data is unique.[11]

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%

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

[11] 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.

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 Amazon.

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.[12] 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.[13] We have estimated the model with different thresholds for the spline and 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 unpopular books.

Store Entry: Our main analysis examines how offline retail store entry influences buyer choice online. Retail store entry will decrease the average distance consumers must travel to offline retailers and also increase 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 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

[12] Due to data constraints, 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.
[13] "Greater than 15,000" is used as the base in the regression analysis.

after a Barnes & Noble or Borders bookstore has entered within a 5.4 mile radius of the location.[14] These data were collected either through press releases from the companies or through direct communication with company representatives. We use 5.4 miles because this is the distance that the average consumer travels to go to a bookstore (Brynjolfsson and Smith 2000).[15] Across our entire sample, 16.4% of locations experience discount store entry, while 4.7% experience a large bookstore entry.

Some locations in our raw data set do not appear in Purchase Circles for the entire sample. To remove concerns that entry and exit of locations in our data set may be somehow correlated with retail store entry, we used only those locations that appear in the sample throughout the 10 months.[16] This resulted in 1501 locations. 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 (Appendix Table A3). 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.

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 propositions based on cross-sectional variation in distance to stores across locations. Empirically, we convert these propositions 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

[14] To compute radii, we use the average longitude and latitude across zip codes within the location.
[15] Results are robust (and in fact stronger) when we use a larger radius of 20 miles. See appendix Table A1.
[16] The number of locations in the sample expanded significantly in November 2005.

location j in month t.[17] 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 regression:

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

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; $\mathbf{NationalRank}_{it}$ 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;[18] μ_{ij} is a product-location fixed effect, μ_t is a time 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.[19]

Our propositions from Section 2 easily convert into testable hypotheses on the coefficients on the interaction of local supply characteristics and product characteristics. Proposition 1 on convenience suggests that decreases in distance to offline stores will be associated with relatively fewer purchases of the most 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). Proposition 2 on

[17] Since our results are primarily based on interaction terms, using a non-linear model such as Probit model would have been harder 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 but given the large number of observations, this becomes less important.

[18] 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.

[19] 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.

selection suggests that entry by large bookstores with larger selection will be associated with fewer purchases of relatively rare or less popular products online relative to entry by discount stores. In particular, the typical large book seller like Barnes & Noble carries between 40,000 and 100,000 books (Brynjolfsson, Hu and Smith 2003) and the typical Wal-Mart carries between 1000-5000 media products, including books, CDs, and DVDs.[20] Therefore, to identify the effects of selection (Proposition 2), we examine two things. First, we examine whether Barnes & Noble entry is negatively associated with purchases of unpopular books (those nationally ranked 5001 to 15,000). Second, we examine whether Barnes & Noble entry has a more negative impact on the popularity of these relatively rare books than Wal-Mart entry. Proposition 3a suggests that as online prices fall relative to offline prices, online sales rise: i.e., $\theta_1 < 0$. Proposition 3b suggests that local discount stores will mitigate this effect because they discount the same types of books as the online retailer: i.e., $\theta_2 > 0$ and $\theta_3 > 0$. The expected signs of the coefficients of the interactions based on these hypotheses are summarized in Table 2.

Thus, 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.[21] 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.[22]

5. Analysis

In this section, we show that changes in local retail options have a substantial effect on the types of products bought online. Table 3 presents our main results. We present estimates for all locations (column 1) as well as separate estimates for small (under 100,000; column 2) and large (over 1 million; column 3)

---

[20] Anderson, C. "Updated Data", www.longtail.com. (July 2005).

[21] Because of the difference-in-difference model, our results are unlikely to be affected by other factors such as online taxes or the number of offline libraries in various locations as long as taxes do not change during the duration of our data collection or there is no correlation between the opening of libraries with that of Wal-Mart, Target, Barnes and Noble, or Borders stores.

[22] Moreover, although in our theory model we argue that online and offline channels will be substitutes for one another, our empirical model also allows us to test the alternative hypothesis that they are complements.

locations.[23] We divide by location size to better understand how local retail supply affects online behavior. In particular, it is well known that the impact of the marginal entrant on local market competition is typically declining in the number of competitors (e.g., Bresnahan and Reiss 1991). Since the number of competitors is likely to be correlated with location size, we expect that the coefficient estimates for our entry variables may vary systematically with the size of the location.

We first examine how offline store entry influences online buyer behavior through changes in *convenience*. As noted above, because of improved convenience, entry by discount stores and large bookstores should decrease purchases of books that are more easily available offline (Proposition 1). This will be particularly true for the most popular products that are available in most discount stores and large bookstores.

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.

In columns (2) and (3), we further examine how these results vary by location size. In small locations (column 2), only entry by discount stores is significant and only for books in the national top 1500. Again, the magnitude is especially large (4.1 percentage points) for the most popular books in the national top 150. It is not surprising that the effect is limited to discount stores in small locations as there is little entry by large bookstores in these locations. Consequently, the test has little power.

Rows 1 through 10 of column (3) show how convenience influences buyer behavior in large locations. These results show that, in contrast to all other types of locations, large bookstore entry in large locations has a significant impact on the likelihood of nationally popular books appearing in a local top 10. In fact, in these large locations, Barnes and Noble and Borders have a bigger impact on convenience than do discount stores across all of our ranges of popularity. Moreover, all of these results are significant at least at the 5% level. Again, the biggest effect is in the national top 150. Large bookstore entry reduces the

[23] The results for locations with populations between 100,000 and 1 million are similar to those for all locations combined and are therefore not shown. They are available from the authors upon request.

likelihood of local top 10 appearances by 3.9 percentage points for these books. Discount store entry reduces the likelihood by 3.0 percentage points for these books. Large bookstore entry shows a statistically significant impact across all books in the rest of the national top 15,000 with marginal effects ranging from 0.1 percentage points to 1.1 percentage points. In sum, in larger locations large bookstores are more important in improving convenience than discount stores.

Perhaps surprisingly, Table 3 provides relatively little evidence of a *selection* effect (Proposition 2). Proposition 2 argues that the marginal effect of store entry over the range of unpopular books will be greater for big stores than for discount stores. Our test of the selection effect relied 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 general, we do not find evidence consistent with the selection hypothesis. The one exception is in large locations, where large bookstore entry decreases the likelihood of a local top 10 appearance by 0.5 percentage points (5% significance), while discount stores have no discernible impact on purchase decisions: these coefficient estimates are significantly different from one another at the 5% level.

Overall, the differences in convenience and selection between small and large locations are informative. An additional discounter has a larger relative impact on a smaller location. The magnitude of the coefficient on the interaction of discounter entry and the most popular books is largest in small towns. An additional retailer makes local book purchases substantially more convenient in these towns. In contrast, we only observe a selection effect in large cities where tastes are more likely to be heterogeneous.

We next examine the effects of a change in *price* on product choice. As expected, an increase in Amazon price relative to offline (list) price decreases the likelihood that a book will be purchased online (Proposition 3a). The coefficient on relative price is negative and statistically significant at the 1% level and is stable across specifications. The coefficient in column (1) (-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 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 Proposition 3a, as its online price decreases, the chances of a book showing up in a local top 10 in the online channel increases.

However, consumers become less sensitive to changes in Amazon price with the entry of discount stores and large bookstores (Proposition 3b). Column (1) 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 marginal effect 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. To put this change into perspective, the one standard deviation change in relative price described above will lead to only a 0.1 percentage point decline in the likelihood of a local top 10 appearance in the presence of retailer entry. Thus, we have significant evidence in support of Proposition 3b: as offline discounts increase due to the entry of new retailers, people are less sensitive to changes in online prices. 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.

This result varies by location size. In small locations, entry by discounters decreases the sensitivity to changes in online price (the coefficient is 0.0215; 5% significance). However, entry by large bookstores has no significant impact on the likelihood of a product showing up in a local top 10 list. As noted above in our discussion of the effects of convenience, this may be in part due to a measurement problem: big stores rarely enter these locations. In large locations, entry by large bookstores had little impact (both in lack of significance and in estimated magnitude) on consumers' sensitivity to price. However, entry by discount stores had a significant and substantial impact: the coefficient estimate on the interaction of discount store entry with relative price is 0.0241, the largest of any size category. This may be because the marginal bookstore entrant has relatively little impact on price sensitivity in larger locations with many stores; however, entry by discounters has a bigger impact because of the relatively smaller number of discounters in these locations (Hudson and McWilliams 2006).

The R-squared of our models range between 0.05 and 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 3 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 more popular products to less popular 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.

5.1 Robustness Checks

In our online appendix Tables A.1 through A.4 we present a number of robustness checks. In Table A.1 we show the results of changing our entry radius from 5.4 to 20 miles. In Table A.2 and A.3 we examine how our results change for different splines of book popularity and for the use of a continuous measure of book popularity rather than a spline. In Table A.4 we explore different choice set definitions.

In Table A.1 we examine the robustness of our results to an entry radius of 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. Moreover many websites report stores only within 20 miles of a customer’s zip code (Chiou 2005). The results are qualitatively similar to those in Table 3. All coefficient estimates on our covariates of interest that were significant in Table 3 retain their same sign and remain significant in Table A.1. Moreover, some coefficients that were insignificant in Table 3 become significant in Table A.1. All of the estimates are in the expected direction. Table A.2 includes a different spline to examine how convenience and selection influence buyer behavior, while Table A.3 uses log(Sales Rank) to measure book popularity. Both sets of models yield qualitatively similar results.

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 purchased only once. 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, displayed in Appendix Table A.4, are qualitatively similar to those in Table 3. In particular, the results on the most popular books and the results on price change little. The main difference is that many of the coefficient estimates for the interaction of the dummies for lower rankings with new store entry are small and statistically insignificant. We believe this is because the fixed effects are inappropriate for the large number of products in these categories for which we have a very small number of observed appearances in the local top 10.

5.2 Further exploration of price, convenience, and selection

Online-offline substitution is likely to vary based on the demographic characteristics of a location. It may also differ based on the characteristics of the product. These differences will have important implications for managers who wish to understand consumer buying behavior. This section extends our previous analysis of how channel substitution varies across densely and sparsely populated areas. We first examine how online-offline substitution varies based on whether a town has a university located in it. Second, we examine online-offline substitution based on the type of book.[24] Third, we examine the impact of taxes.

Table 4 presents these results. Columns (1) and (2) show how our results vary based on whether the town has a university located in it. University towns may have a greater number of consumers that have tastes for less popular books, thus the substitution between online and offline channels may be different in university towns than in other types of locations. 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

[24] We also explored how such substitution varies for a variety of other demographic characteristics, including percent of population with a college degree, per capita income, population age distribution, and broadband Internet penetration. In general, we found relatively small differences in channel substitution across these demographic groups. Because the purpose of this section is to reveal differences in substitution behavior that may be important to managers and because of space constraints, we discuss only the differences based on university presence and book characteristics.

substitute away from the most popular books.

However, new store entry in university towns also causes consumers to substitute away from less popular books. For university towns the coefficients on the interaction of new store entry with lower ranked dummies are all statistically significant at the 1% level, ranging in size from -0.0023 to -0.0141. In contrast, for locations without a university only two of the lower-ranked interaction terms are significant. The university town results also fulfill the second (main) criterion for our selection proposition (Proposition 2): the coefficient on large book stores is more negative than that for discounters for relatively unpopular books. For the coefficients on the interaction between the national top 5001-15,000 books, these coefficients are significantly different from one another at the 5% level. Further, the coefficients on the interaction between the national top 1501-5000 books are significantly different at the 1% level. No such differences are evident in non-university towns. In short, it appears that consumers in university towns (where consumers may prefer relatively less popular books) use online retailers to achieve better selection, while consumers in towns without a university do not. When relatively rare books become available offline in these locations, purchase patterns change enough that we observe a fall in the online popularity of these books.

In addition to examining how our results vary by type of location, we also examine how they vary by type of product. Columns (3) and (4) of Table 4 show how our results vary for fiction and non-fiction books. The parameter estimates 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 than those in column (1) in Table 3 (and for non-fiction books). In contrast, many of these estimates for non-fiction books are small and statistically insignificant. However, the effects of price are qualitatively similar for fiction and nonfiction books.[25]

In short, the importance of the online channel for convenience is driven primarily by fiction books. We speculate that this result is driven by non-fiction purchases being more directed than fiction purchases. The expected utility to an individual from reading a particular fiction book may be less certain than the

[25] One potential concern with this analysis is that the propensity to purchase fiction books may change over time in our sample, shifting the relative percentage of fiction and nonfiction books in our choice set. We re-estimated the model in Table 4 using a choice set that had a constant fraction of fiction and nonfiction books. The results were qualitatively similar.

expected utility from reading a particular non-fiction book. Mehta, Rajiv, and Srinivasan (2003) show that a larger variance in consumption utility leads to a larger consideration set. If more fiction books are under consideration, the small set of books available at Wal-Mart (and on prominent display at Barnes and Noble) are more likely to be in a customer's consideration set if they are fiction. This provides a likely explanation for the result that entry has a larger impact on fiction than non-fiction.

We also examined how our results varied if sales taxes were assessed for online purchases in a location. Sales taxes are assessed for Amazon purchases in only four states: Kansas, Kentucky, North Dakota, and Washington. Prior work has demonstrated that the assessment of online sales taxes influences consumer propensity to buy online (Goolsbee 2000; Ellison and Ellison 2006). Though it is not our primary goal to assess the impact of sales taxes on electronic commerce, we examine how online-offline channel substitution varies with online sales taxes to further validate our findings and those of prior work on taxes and electronic commerce. 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 sales tax states than in others. That is, entry by offline retailers in sales tax states will increase the likelihood that consumers will substitute an offline purchase for an online purchase, relative to entry by offline retailers in states without sales taxes. The results in columns (5) and (6) of Table 4 are consistent with this assertion.

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), improved 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 is likely to vary significantly across locations in the physical world. 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 bridging empirical electronic commerce research with related literatures in marketing and economics on buyer behavior and product market competition in the offline world. Table 5 summarizes our main results.

Our emphasis on the roles of convenience and selection in online buyer behavior advances a literature on online-offline channel substitution that has thus far focused primarily on the ability of the online channel to offer lower prices. In fact, our results showed that convenience was the single biggest factor influencing online-offline substitution. This suggests that future work on channel substitution should examine the role of convenience on online buying behavior. More broadly, though prior work has shown that convenience may influence the dollar value of spending on internet commerce, to our knowledge no prior work has examined how convenience shapes consumers' product choices online. Our results suggest that this should be an area of continuing research.

We find the impact of selection is relatively small, and is isolated locations such as 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; Brynjolfsson, Hu, and Smith 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 for books. To put it another way, offline bookstores are better than online retailers for browsing when consumers are unsure of which product to buy, but online retailers are much better when consumers know what they wish to buy. Our results on fiction books support this possibility. That is to say, browsers who use large offline stores may represent a slightly different market than online users, so there is limited substitution between the two.

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. For example, to the extent that offline entry is correlated with a shift away from purchasing the most popular products online, electronic markets may want to shift toward targeting less popular products. 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 most popular 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 influence 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.

We present a theoretical model of consumer behavior and then examine the extent to which our data are consistent with this model. However, 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 are consistent with alternative models of the mechanisms driving this substitution, 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) or with more non-digital attributes (Lal and Sarvary 1999) that consumers may prefer to experience before purchasing online.

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 urban customers.

We develop a theoretical model that examines why local retail options might influence the types of products bought online. This model demonstrates how variance in convenience, selection, and price across local markets may influence consumer behavior online. Our empirical results support the propositions in this model. In particular, we demonstrate that offline convenience and selection play a significant role in driving consumer decisions to use the online or offline channel.

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.

- Balasubramaniam, S. 1998. Mail versus mall: a strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science* 17(3), 181-195.

- 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.

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

- 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.

- Brynjolfsson, E., Y. Hu, and M. Smith. 2006. From Niches to Riches: Anatomy of the Long Tail. *MIT Sloan Management Review* 47(4), 67-71.

- 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.

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

- 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.

- Emre, O., A. Hortacsu, and C. Syverson. 2005. E-commerce and the Market Structure of Retail Industries. NET Institute Working Paper #05-24.

- 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.

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

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

- Granados, N, A. Gupta, and R. Kauffman. 2006. IT-Enabled Transparent Electronic Markets: The Case of The Air Travel Industry," *Forthcoming in Information Systems and e-Business Management.*

- Hortacsu, A., F.A. Martínez-Jerez, and J. Douglas. 2006. The Geography of Trade on eBay and MercadoLibre. NET Institute Working Paper #06-09.

- Hudson, K. and G. McWilliams. 2006. Seeking Growth in Urban Areas, Wal-Mart Gets Cold Shoulder. *The Wall Street Journal,* September 25, 2006, A1 and A8.

- 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.

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

- Lee, T. and I. Png 2004. Buyer Switching Costs and Retail Pricing: An Indirect Empirical Test, *Review of Marketing Science*, 2 (6), 1-22.

- Mehta, N., S. Rajiv, and K. Srinivasan. 2003. Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science* 22(1), 58-84.

- Owen, T. 2002. Lord of the things. Business 2.0 (March) 32.
- Prince, J. 2006. The Beginning of Online/Retail Competition and its Origins: An Application to Personal Computers. Forthcoming, *International Journal of Industrial Organization.*
- 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.
- Zhu, T., A. Dukes and V. Singh. 2005. Local Competition and Entry by a Discount Retailer, Working Paper.

Appendix

Proof of Proposition 1

From equations (5) and (6), we can derive the profit functions for the two firms. Note that there are six unknown parameters-λ_1, λ_2, p_{A1}, p_{B1}, p_{A2}, and p_{B2}. This gives us six reaction functions that are derived by setting the six first order conditions to zero. The solutions to these reactions functions gives us the optimal parameter values for λ^*_1, λ^*_2, p^*_{A1}, p^*_{B1}, p^*_{A2}, and p^*_{B2}.[26]

As mentioned before, the equilibrium values of realized demand for the unpopular and popular product for the online firm are given by M_{B1} and M_{B2}, respectively. The expression $\frac{M_{B2}}{M_{B1}+M_{B2}}$ therefore denotes the proportion of popular products that is being bought online. Taking the first order condition of the expression $\frac{M_{B2}}{M_{B1}+M_{B2}}$ with respect to d, yields the desired results stated in the Proposition. Specifically, we find that

$$\frac{\partial}{\partial d}\left(\frac{M_{B2}}{M_{B1}+M_{B2}}\right)=\left[\frac{(1-r)r(r-\Phi)(b_1-b_2)b_2C}{\left(2(1-r)r(d(b_1-b_2)+b_2)+(r+\Phi-2r\Phi)Cb_2\right)^2}\right]$$

In equilibrium this is positive for all r > Φ. This proves the Proposition.[27]

Proof of Proposition 2

The equilibrium values of realized demand for the unpopular and popular product for the online firm are given by M_{B1} and M_{B2}, respectively. The expression $\frac{M_{B1}}{M_{B1}+M_{B2}}$ therefore denotes the proportion of less popular products that is being bought online. Taking the first order condition of the expression $\frac{M_{B1}}{M_{B1}+M_{B2}}$ with respect to Φ, yields the desired results stated in the Proposition. Specifically, we find that

$$\frac{\partial}{\partial \Phi}\left(\frac{M_{B1}}{M_{B1}+M_{B2}}\right)==\left[-\frac{b_2(1-r)r(d(b_1-b_2)+b_2(1-C))C}{\left(2(1-r)r(d(b_1-b_2)+b_2)+(r+\Phi-2r\Phi)Cb_2\right)^2}\right]$$

In equilibrium this expression is negative since $b_1 > b_2$.[28]

[26] Details of the expressions for optimal prices and profit functions are omitted for brevity but are available from the authors upon request.

[27] Note that the qualitative nature of this Proposition holds true even if we were to denote the fraction of popular products being bought online as $\frac{M_{B2}}{M_{A2}+M_{B2}}$. We can prove that $\frac{\partial}{\partial d}\left(\frac{M_{B2}}{M_{A2}+M_{B2}}\right)>0$

Proof of Proposition 3a

With a discount of ω_{A2} and ω_{B2} on popular books across the offline and online channels respectively, the demand function for popular books sold through the offline channel is given by

$$(1-d-\omega_{A2}b_1p_{A2} + \omega_{B2}b_2p_{B2})\ \omega_{A2}p_{A2}$$

Similarly, the demand function for popular books sold through the online channel is given by

$$(d - \omega_{A2}b_1p_{B2} + \omega_{B2}b_2p_{A2})\ \omega_{A2}p_{A2}$$

Hence, the optimal prices for popular books are given by

$$p_{A2} = \frac{(-d(2b_1 - b_2) + 2b_1(1 - C + \Phi C))}{(2b^2{}_1 - b_2{}^2)\omega_{A2}}$$

$$p_{B2} = \frac{(d(b_1 - b_2) + b_2(1 - C + \Phi C))}{(2b_1^2 - b_2^2)\omega_{B2}}$$

From the above equation it is immediate that $\frac{\partial p_{B2}}{\partial \omega_{B2}} < 0$ and hence it follows that $\frac{\partial M_{B2}}{\partial \omega_{B2}} > 0$ because of the fact that the own price effect is negative. Thus, with an increase in online discounts for popular books, there is an increase in sales of the popular books through the online channel.

Proof of Proposition 3b

Consider the effect of a decrease in distance to the offline stores. This is captured in our model by a decrease in *d*. Then differentiating $\frac{\partial p_{B2}}{\partial \omega_{B2}}$ with respect to *d*, we find that $\frac{\partial^2 p_{B2}}{\partial d \partial \omega_{B2}} = -\frac{b_1 - b_2}{(2b_1^2 - b_2^2)\omega^2{}_{B2}}$ Hence, $\frac{\partial^2 p_{B2}}{\partial d \partial \omega_{B2}} < 0$. This implies that $\frac{\partial^2 M_{B2}}{\partial d \partial \omega_{B2}} > 0$. That is, an increase in online sales due to an increase in online price discounts is mitigated by the reduction in *d*, i.e. by an increase in the entry of retail stores such as Wal-Mart and Barnes & Noble. This completes the proof.[29]

---

[28] This result holds true even if we were to denote the fraction of popular products being bought online as $\frac{M_{B1}}{M_{A1}+M_{B1}}$.
We can prove that $\frac{\partial}{\partial \Phi}\left(\frac{M_{B1}}{M_{A1}+M_{B1}}\right) < 0$

[29] We assume in the model that the conditions needed for prices to be positive are satisfied, i.e., the required relationships between the exogenous parameters hold.

Table 1: Summary Statistics for Books

Variable	Observations	Mean	Std. Dev.	Min	Max
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 Hypotheses

Variable	*Expected Sign*	*Intuition*
Convenience (Hypothesis 1)		
(Very popular books)* (Discounter entry)	Negative	With an increase in the number of discount stores, more popular products are bought offline
(Very popular books)* (Large bookstore entry)	Negative	With an increase in the number of large bookstores, more popular products are bought offline
Selection (Hypothesis 2)		
(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.
Price (Hypothesis 3)		
Relative price	Negative	As the Amazon price (relative to the list price) increases, the product is less likely to be purchased online, other things equal.
(Relative price)* (Discounter entry)	Positive	With an increase in the number of discount stores, consumers will be less sensitive to changes in online price.
(Relative price)* (Large bookstore entry)	Positive	With an increase in the number of large bookstores, consumers will be less sensitive to changes in online price.

Table 3: Main Book Results—Difference in Difference on store entry

	(1)	(2)	(3)
	All Data	Locations Under 100k	Locations over 1 million
(Top 150 Books)* (Discount Store Entry within 5.4 miles)	-0.0320	-0.0409	-0.0299
	(0.0012)**	(0.0045)**	(0.0023)**
(Top 150 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0339	-0.0067	-0.0389
	(0.0025)**	(0.0176)	(0.0045)**
(Top 151-500 Books)* (Discount Store Entry within 5.4 miles)	-0.0034	-0.0071	0.0014
	(0.0008)**	(0.0030)*	(0.0016)
(Top 151-500 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0029	-0.0025	-0.0073
	(0.0020)	(0.0142)	(0.0034)*
(Top 501-1500 Books)* (Discount Store Entry within 5.4 miles)	-0.0060	-0.0065	-0.0024
	(0.0006)**	(0.0028)*	(0.0013)+
(Top 501-1500 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0022	0.0112	-0.0061
	(0.0016)	(0.0115)	(0.0026)*
(Top 1501-5000 Books)* (Discount Store Entry within 5.4 miles)	-0.0082	-0.0072	-0.0088
	(0.0009)**	(0.0050)	(0.0019)**
(Top 1501-5000 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0074	0.0172	-0.0105
	(0.0025)**	(0.0116)	(0.0040)**
(Top 5001-15,000 Books)* (Discount Store Entry within 5.4 miles)	-0.0019	-0.0006	0.0014
	(0.0007)**	(0.0037)	(0.0016)
(Top 5001-15,000 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0022	0.0167	-0.0049
	(0.0018)	(0.0157)	(0.0024)*
Relative price	-0.0237	-0.0221	-0.0275
	(0.0007)**	(0.0023)**	(0.0015)**
(Relative Price)* (Discount Store Entry within 5.4 miles)	0.0147	0.0215	0.0241
	(0.0022)**	(0.0102)*	(0.0046)**
(Relative Price)* (Large Bookstore Entry within 5.4 miles)	0.0183	0.0160	0.0041
	(0.0061)**	(0.0196)	(0.0090)
Top 150 Books	0.006	0.0050	0.0066
	(0.0004)**	(0.0014)**	(0.0008)**
Top 151-500 Books	-0.0124	-0.0104	-0.0135
	(0.0003)**	(0.0010)**	(0.0006)**
Top 501-1500 Books	-0.0061	-0.0041	-0.0069
	(0.0003)**	(0.0009)**	(0.0005)**
Top 1501-5000 Books	-0.0066	-0.0047	-0.0077
	(0.0002)**	(0.0007)**	(0.0004)**
Top 5001-15,000 Books	-0.0041	-0.0023	-0.0046
	(0.0001)**	(0.0004)**	(0.0002)**
Observations	4,062,326	386,551	1,119,715
Number of FEs	981,255	93,393	271,419
R-squared	0.0580	0.0535	0.0588
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 -Log(Number of reviews) -Discount Store Entry within 5.4 miles - Large Bookstore Entry within 5.4 miles		

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

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

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 3.

+ significant at 10%; * significant at 5%; ** significant at 1%

Table 5: Summary of results

Hypothesis ***(Derived from the Propositions in section 3)***	***Data Used***						
	All Data	**Larger Locations**	**Locations with Universities**	**Locations without Universities**	**Fiction Books Only**	**Books that are not fiction**	**Locations with taxes**
Table and column	**Table 3 column 1**	**Table 3 column 3**	**Table 4 column 1**	**Table 4 column 2**	**Table 4 column 3**	**Table 4 column 4**	**Table 4 column 6**
Hypothesis 1: Convenience ***-coefficient on interaction of entry and most popular books is negative***	Strongly supported	Strongly supported	Strongly supported	Strongly supported	Strongly supported & much higher than for other books	Weakly supported for discounters	Strongly supported
Hypothesis 2: Selection ***-coefficient on interaction of large store entry and less popular books is negative, and less than the same coefficient for discounters***	Not supported	Supported	Supported	Not supported	Not supported	Not supported	Supported
Hypothesis 3a: Online price ***-coefficient on relative price is negative***	Strongly supported	Strongly supported	Strongly supported	Strongly supported	Strongly supported	Strongly supported	Strongly supported
Hypothesis 3b: Offline price ***-coefficient on interaction of entry and relative price is positive***	Strongly supported	Strongly supported for discounters	Strongly supported	Strongly supported	Strongly supported & much higher than for other books	Strongly supported	Strongly supported

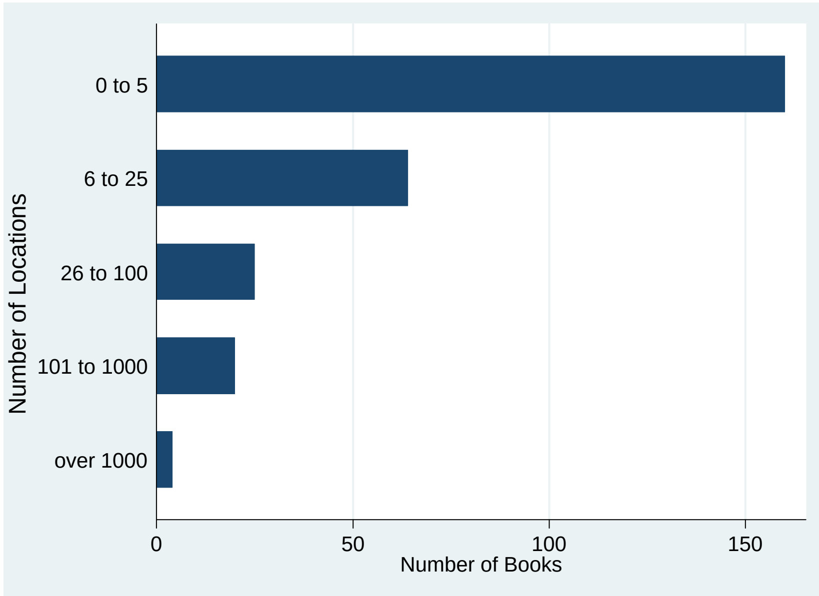

Figure 1: Number of Locations that a Book is in the Top 10--May 2005

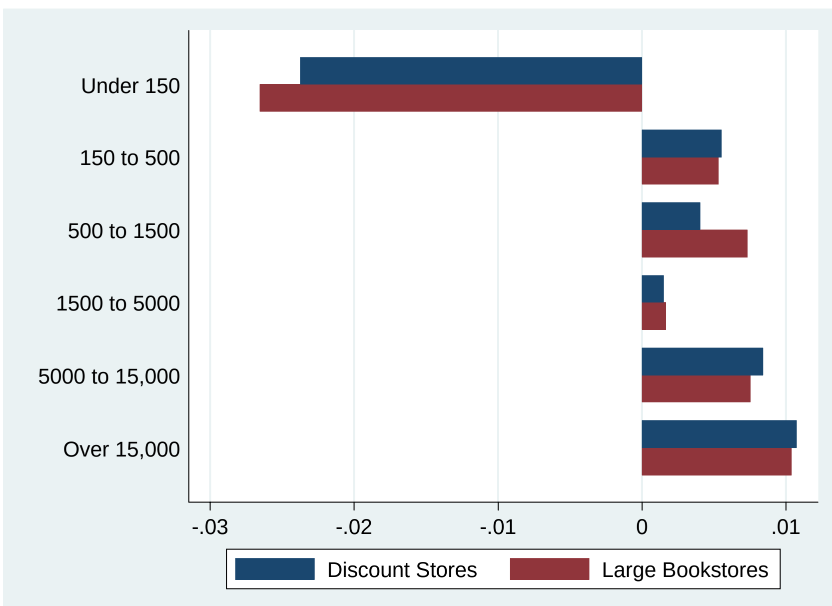

Figure 2: Discount Store/Specialty Store Comparison
Marginal Effect of Store Entry by Sales Rank,
(Based on Table 3 column (1))

Technical Appendix

Appendix Table A1: Robustness Check—Uses distance to store of 20 miles

	(1)	(2)	(3)
	All Data	Locations Under 100k	Locations over 1 million
(Top 150 Books)* (Discount Store Entry within 20 miles)	-0.0372	-0.0384	-0.0433
	(0.0008)**	(0.0033)**	(0.0014)**
(Top 150 Books)* (Large Bookstore Entry within 20 miles)	-0.0343	-0.0179	-0.0349
	(0.0011)**	(0.0133)	(0.0020)**
(Top 151-500 Books)* (Discount Store Entry within 20 miles)	-0.0061	-0.0064	-0.0076
	(0.0005)**	(0.0022)**	(0.0008)**
(Top 151-500 Books)* (Large Bookstore Entry within 20 miles)	-0.0045	0.0023	-0.0014
	(0.0008)**	(0.0108)	(0.0015)
(Top 501-1500 Books)* (Discount Store Entry within 20 miles)	-0.0080	-0.0059	-0.0088
	(0.0003)**	(0.0017)**	(0.0006)**
(Top 501-1500 Books)* (Large Bookstore Entry within 20 miles)	-0.0048	0.0086	-0.0047
	(0.0006)**	(0.0098)	(0.0012)**
(Top 1501-5000 Books)* (Discount Store Entry within 20 miles)	-0.0084	-0.0064	-0.0116
	(0.0004)**	(0.0027)*	(0.0008)**
(Top 1501-5000 Books)* (Large Bookstore Entry within 20 miles)	-0.0067	0.0096	-0.0082
	(0.0009)**	(0.0127)	(0.0017)**
(Top 5001-15,000 Books)* (Discount Store Entry within 20 miles)	-0.0020	-0.0013	-0.0013
	(0.0003)**	(0.0018)	(0.0005)*
(Top 5001-15,000 Books)* (Large Bookstore Entry within 20 miles)	-0.0021	0.0113	-0.0035
	(0.0006)**	(0.0147)	(0.0011)**
Relative price	-0.0268	-0.0229	-0.0309
	(0.0008)**	(0.0023)**	(0.0016)**
(Relative Price)* (Discount Store Entry within 20 miles)	0.0107	0.0177	0.0106
	(0.0010)**	(0.0060)**	(0.0018)**
(Relative Price)* (Large Bookstore Entry within 20 miles)	0.0145	0.0057	0.0130
	(0.0021)**	(0.0176)	(0.0038)**
Top 150 Books	0.0180	0.0082	0.0241
	(0.0005)**	(0.0015)**	(0.0010)**
Top 151-500 Books	-0.0108	-0.0101	-0.0115
	(0.0003)**	(0.0010)**	(0.0006)**
Top 501-1500 Books	-0.0042	-0.0038	-0.0044
	(0.0003)**	(0.0009)**	(0.0005)**
Top 1501-5000 Books	-0.0045	-0.0043	-0.0043
	(0.0002)**	(0.0007)**	(0.0004)**
Top 5001-15,000 Books	-0.0033	-0.0021	-0.0034
	(0.0001)**	(0.0005)**	(0.0002)**
Observations	4,062,326	386,551	1,119,715
Number of FEs	981,255	93,393	271,419
R-squared	0.0582	0.0536	0.0591
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 -Log(Number of reviews) -Discount Store Entry within 20 miles -Large Bookstore Entry within 20 miles		

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%

Appendix Table A2: Robustness Check—Uses a different spline for book sales rank

	(1)	(2)	(3)
	All Data	Locations Under 100k	Locations over 1 million
(Top 100 Books)* (Discount Store Entry within 5.4 miles)	-0.0370	-0.0450	-0.0370
	(0.0014)**	(0.0055)**	(0.0026)**
(Top 100 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0360	-0.0093	-0.0388
	(0.0027)**	(0.0148)	(0.0048)**
(Top 101-500 Books)* (Discount Store Entry within 5.4 miles)	-0.0061	-0.0126	-0.0017
	(0.0008)**	(0.0030)*	(0.0016)
(Top 101-500 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0058	-0.0040	-0.0092
	(0.002) **	(0.0145)	(0.0036)*
(Top 501-1000 Books)* (Discount Store Entry within 5.4 miles)	-0.0053	-0.0087	-0.0028
	(0.0005)**	(0.0028)**	(0.0012)+
(Top 501-1000 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0010	0.0120	-0.0040
	(0.0016)	(0.0110)	(0.0026)
(Top 1001-10000 Books)* (Discount Store Entry within 5.4 miles)	-0.0060	-0.0060	-0.0060
	(0.0006)**	(0.0030)**	(0.0015)**
(Top 1001-10000 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0036	0.0120	-0.0050
	(0.0018)*	(0.0090)	(0.0027)**
Relative price	-0.0165	-0.0161	-0.0202
	(0.0007)**	(0.0022)**	(0.0014)**
(Relative Price)* (Discount Store Entry within 5.4 miles)	0.0124	0.0199	0.0210
	(0.0021)**	(0.0102)*	(0.0045)**
(Relative Price)* (Large Bookstore Entry within 5.4 miles)	0.0170	0.0145	0.0037
	(0.0061)**	(0.0200)	(0.0090)
Top 100 Books	0.0410	0.0360	0.0042
	(0.0005)**	(0.0016)**	(0.001)**
Top 101-500 Books	-0.0090	-0.0090	-0.0100
	(0.0002)**	(0.0007)**	(0.0004)**
Top 501-1000 Books	-0.0030	-0.0025	-0.0033
	(0.0001)**	(0.0005)**	(0.0003)**
Top 1001-10000 Books	-0.0023	-0.0017	-0.0027
	(0.0001)**	(0.0004)**	(0.0002)**
Observations	4,062,326	386,551	1,119,715
Number of FEs	981,255	93,393	271,419
R-squared	0.0606	0.0558	0.0615
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 -Log(Number of reviews) -Discount Store Entry within 5.4 miles -Large Bookstore Entry within 5.4 miles		

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%

Appendix Table A3: Robustness Check—Uses log(Sales Rank) instead of spline for books sales rank

	(1)	(2)	(3)
	All Data	Locations Under 100k	Locations over 1 million
Log(Sales Rank)* (Discount Store Entry within 5.4 miles)	0.0034	0.0049	0.0031
	(0.0002)**	(0.0007)**	(0.0004)**
Log(Sales Rank)* (Large Bookstore Entry within 5.4 miles)	0.0035	0.0001	0.0039
	(0.0004)**	(0.0023)	(0.0008)**
Relative price	-0.0146	-0.0138	-0.0194
	(0.0007)**	(0.0023)**	(0.0015)**
(Relative Price)* (Discount Store Entry within 5.4 miles)	0.0153	0.0206	0.0265
	(0.0022)**	(0.0106)*	(0.0047)**
(Relative Price)* (Large Bookstore Entry within 5.4 miles)	0.0229	0.0174	0.0081
	(0.0063)**	(0.0228)	(0.0092)
Log(Sales Rank)	-0.0100	-0.0097	-0.0091
	(0.0001)**	(0.0004)**	(0.0003)**
Observations	4,062,326	386,551	1,119,715
Number of FEs	981,255	93,393	271,419
R-squared	0.0602	0.0556	0.0608
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 -Log(Number of reviews) -Discount Store Entry within 5.4 miles -Large Bookstore Entry within 5.4 miles		

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%

Appendix Table A4: Robustness Check—Uses a larger set of books in the choice set

	(1)	(2)	(3)
	All Data	Locations Under 100k	Locations over 1 million
(Top 150 Books)* (Discount Store Entry within 5.4 miles)	-0.0343	-0.0412	-0.0373
	(0.0011)**	(0.0044)**	(0.0020)**
(Top 150 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0387	-0.0173	-0.0424
	(0.0023)**	(0.0141)	(0.0039)**
(Top 151-500 Books)* (Discount Store Entry within 5.4 miles)	-0.0018	-0.0032	-0.0004
	(0.0005)**	(0.0018)+	(0.0009)
(Top 151-500 Books)* (Large Bookstore Entry within 5.4 miles)	-0.0017	-0.0068	-0.0045
	(0.0012)	(0.0075)	(0.0020) *
(Top 501-1500 Books)* (Discount Store Entry within 5.4 miles)	0.0018	0.0016	0.0043
	(0.0003)**	(0.0012)	(0.0005)**
(Top 501-1500 Books)* (Large Bookstore Entry within 5.4 miles)	0.0041	0.0053	0.0019
	(0.0007)**	(0.0045)	(0.0011)+
(Top 1501-5000 Books)* (Discount Store Entry within 5.4 miles)	-0.0007	-0.0003	0.00002
	(0.0003)+	(0.0019)	(0.0007)
(Top 1501-5000 Books)* (Large Bookstore Entry within 5.4 miles)	0.0005	0.0077	-0.0007
	(0.0009)	(0.0036)*	(0.0013)
(Top 5001-15,000 Books)* (Discount Store Entry within 5.4 miles)	0.0006	0.0007	0.002
	(0.0002)*	(0.0014)	(0.0004)**
(Top 5001-15,000 Books)* (Large Bookstore Entry within 5.4 miles)	0.0009	0.0062	0.0003
	(0.0005)+	(0.0044)	(0.0007)
Relative price	-0.0149	-0.0134	-0.0177
	(0.0004)**	(0.0012)**	(0.0008)**
(Relative Price)* (Discount Store Entry within 5.4 miles)	0.0166	0.0193	0.0242
	(0.0012)**	(0.0054)**	(0.0026)**
(Relative Price)* (Large Bookstore Entry within 5.4 miles)	0.0207	0.015	0.0144
	(0.0031)**	(0.0128)	(0.0046)**
Top 150 Books	0.0141	0.0116	0.0161
	(0.0003)**	(0.0010)**	(0.0006)**
Top 151-500 Books	-0.0075	-0.0068	-0.0077
	(0.0002)**	(0.0006)**	(0.0003)**
Top 501-1500 Books	-0.0011	-0.0002	-0.0014
	(0.0001)**	(0.0004)	(0.0002)**
Top 1501-5000 Books	-0.0007	0.00004	-0.0011
	(0.0001)**	(0.0003)	(0.0002)**
Top 5001-15,000 Books	-0.0002	0.0004	-0.0004
	(0.0001)**	(0.0002)*	(0.0001)**
Observations	9,446,406	899,538	2,602,576
Number of Fes	2,941,782	280,574	812,601
R-squared	0.0355	0.0321	0.0365
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 -Log(Number of reviews) -Discount Store Entry within 5.4 miles -Large Bookstore Entry within 5.4 miles		

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%.