Competition between Local and Electronic Markets: How the benefit of buying online depends on where you live*

Chris Forman
College of Management
Georgia Institute of Technology
800 West Peachtree Street NW
Atlanta, GA 30308
chris.forman@mgt.gatech.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

September 2007

Abstract

We empirically examine the trade-off between the benefits of buying online and the benefits of buying in a local retail store. How does a consumer's physical location shape the relative benefits of buying from the online world? We explore this problem using data from Amazon on the top selling books for 1497 unique locations in the US for 10 months ending in January 2006. In particular, we examine what happens when a large bookstore opens and when a discount retailer opens. We show that even controlling for product-specific preferences by location, changes in local retail options have substantial effects on online purchases. When a store opens locally, we find evidence that people substitute away from online purchasing, demonstrating that consumers appear to respond to increased convenience in the offline channel. These estimates are economically large, suggesting that disutility costs of purchasing online are substantial and that offline transportation costs matter. We also show that offline entry decreases consumers' sensitivity to online price discounts. We find no consistent evidence that the breadth of the product line at a local retail store affects purchases although breadth seems to matter in university towns and larger cities. Our paper shows that the parameters in existing theoretical models of channel substitution such as offline transportation cost, online disutility cost, market coverage, and the prices of online and offline retailers interact to determine consumer choice of channels. In this way, our results provide empirical support for many such models.

* We thank participants at the Second Annual Symposium on Statistical Challenges in E-Commerce, the Marketing Science 2006 Conference, the INFORMS 2006 Conference, the 2006 International Conference in Information Systems, the 2006 Harvard-Wharton Consortium on Operational Excellence in Retailing, the HICSS 20th Anniversary Symposium on Competitive Strategy, Economics, and Information Systems, the Federal Trade Commission, , seminar participants at Carnegie Mellon University, University of Florida, Georgia Tech, University of Toronto, Michigan State University, University of Texas Austin, New York University, University of South Carolina, and three anonymous reviewers of the International Conference in Information Systems (ICIS 2006) for useful feedback. We also thank Jeremy Fox, Korhan Gurkan, Ke-Wei Huang, Steven Klepper, Barrie Nault, Paul Pavlou, and Jeffrey Prince for valuable suggestions. We thank Rong Zheng and Ashley Tyrrel for outstanding assistance with the collection of data, and Gordon Eiland, VP of Strategy at Borders for supplying us with the store opening data of Borders. Support was provided by a Berkman Faculty Development grant, the NET Institute, an NSF CAREER award IIS-0643847 and SSHRC grant # 538-02-1013. All errors are ours alone.

1. Introduction

As of 2006, electronic commerce represented just 3% of total retail sales (US Census 2007). Online shopping remains a small fraction of retail sales despite the well-known benefits of electronic commerce to consumers, including lower prices (e.g. Brynjolfsson and Smith 2000), greater selection (e.g., Brynjolfsson, Hu, and Smith 2003; Ghose, Smith and Telang 2006), and greater convenience by eliminating travel costs and enabling 24x7 purchases irrespective of geographic location (Cairncross 1997). Of course, there are many reasons why consumers do not take advantage of the online channel: Inspecting non-digital products is often difficult, shipping can be time consuming and expensive, and returning products can be challenging. That is, there appear to be a set of fixed disutility costs of buying online. These costs vary significantly across products and retailers, and in some markets have created significant hurdles to the continued diffusion of electronic commerce.

Theoretical research has explored consumers' choice of channel in commodity markets, modeling the decision as a tradeoff between these fixed disutility costs and the lower search and transportation costs of buying online, in addition to any price differences across the two channels (Balasubramanian 1998; Chun and Kim 2005; Liu, Gupta, and Zhang 2006; Cheng and Nault 2007). However, while prior research has demonstrated that consumers have significant cross-price elasticity between online and offline markets (Goolsbee 2000; Ellison and Ellison 2006; Prince 2007), there exists little systematic empirical evidence on the tradeoff between offline transportation costs and online disutility costs. In short, while theory often assumes that the benefits of buying online depend on where you live, we have little evidence on how much it matters.

In this paper, we take one step towards assessing how tradeoffs between online disutility costs and offline transportation costs influence consumer online purchase behavior for commodity products. Motivated by a wide array of theory literature on spatial competition and online-offline competition, we develop a set of hypotheses that examine how changes in offline transportation costs, online and offline prices, consumer reservation values, and online disutility costs influence channel choice. In this way, our paper is in the spirit of prior theory-testing research that has provided empirical validation to theories on how the Internet influences buyer decisions due to lower search costs (Brynjolfsson and Smith 2000), greater product selection (Brynjolfsson, Hu, and Smith 2003), and through better product information due to improved word of mouth (Chevalier and Mayzlin 2006).

We test these hypotheses using monthly data from Amazon.com on top-selling books in 1497 local markets from April 2005 to January 2006. In particular, we examine how entry by retailers Wal-Mart, Target, Barnes & Noble, and Borders changes the types of products bought online in the location where the store entered and compare this to the types of products bought in locations that did not experience such entry. Our method controls for differences in consumer preferences across locations through product-location fixed effects. In this way, we use store entry to identify the effects of improved offline options on online choice using a difference-in-difference strategy. By focusing on books, we study a commodity product where brand-specific and product-specific factors are less likely to influence channel substitution, and where purchase-related factors that cannot be determined digitally are relatively unimportant. Moreover, we study an environment where e-commerce sales are high (so the tradeoff we explore is economically important) and where online disutility costs are likely to be relatively low (so our estimates of the magnitude of online disutility costs relative to offline transportation costs are likely to be conservative, in comparison to other products and industries).

We find evidence that people substitute away from online purchasing toward offline purchasing when a store opens locally: people appear to respond to increased convenience in the offline channel. 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 relative to purchases of products unlikely to be prominent, or even available, offline. We find these effects to be economically large, suggesting that disutility costs of purchasing online are significant even in the book market. We also show that offline entry significantly decreases consumers' sensitivity to online price discounts. However, we do not find consistent evidence that the breadth of the product line at a local retail store affects purchases. Although Barnes & Noble has a much wider selection of books than Wal-Mart, entry by either has the same primary effect: the most popular products become less likely to be bought online. We attribute this in part to (1) high offline transportation costs (in expectation) due to uncertain availability of less popular books at offline stores and (2) limited consumer demand for less popular products. However, we do find evidence that product selection matters in university towns and larger cities, where consumer tastes may be more varied and therefore the concentration of consumers with preferences for less popular products is likely to be greater.

Our paper contributes to three areas of research. First, by motivating our hypotheses based on the findings from a significant body of literature that uses models of spatial competition to examine online-offline channel substitution (Balasubramanian 1998; Cheng and Nault 2007; Chun and Kim 2005; Liu, Gupta, and Zhang 2006; Viswanathan 2005), we provide empirical support for assumptions widely used in theoretical research. By shedding light on the relative magnitudes of some of the parameters used in these models, we provide further insights into results in these papers that often depend on these parameter values.

Second, our paper also contributes to a small empirical literature on consumer substitution between online and offline channels (see for example Goolsbee 2001; Ellison and Ellison 2006; Prince 2006). In contrast to this prior work which focuses on cross price elasticities, our paper explores how relative prices, online convenience, offline product breadth, and online purchasing costs such as the value of offline immediacy influence consumer behavior. While Brynjolfsson, Hu, and Raman (2007) examine the role of local characteristics in women's clothing, they focus on how equilibrium market conditions relate to online choices in a cross section. In contrast, the panel nature of our data means that we can separately identify local demand-side preferences from supply-side factors related to retail competition. Moreover, as mentioned above, our analysis of a commodity product (books), allows us to test assumptions previously used in models of spatial competition.

Third, and more broadly, this paper advances the emerging stream of empirical literature that studies how online and catalog retailing contribute to consumer welfare. One stream of this literature has studied how Internet retailing has influenced price competition and price dispersion, and demonstrated that consumers benefit from lower prices in the online channel (for a review, see Baye, Morgan, and

Scholten 2006). A related line of research has shown that by lowering search costs, Internet retailing improves consumer welfare by helping consumers obtain hard-to-find books (Brynjolfsson, Hu, and Smith 2003), increasing the resale value of new products (Ghose, Telang, and Krishnan 2005), and facilitating the market for used books (Ghose, Smith, and Telang 2006). We contribute to this literature by examining the benefits of Internet retailing in improving customer convenience. In contrast to research that has examined how Internet technology reduces the costs associated with geographic isolation (Forman, Goldfarb, and Greenstein 2005; Sinai and Waldfogel 2004) and improves the information convenience in online trading markets (Balasubramanian, Konana, and Menon 2003), we emphasize how Internet retailing provides a convenient substitute to local retailing when there are transportation costs.

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

2. Hypotheses

Our hypotheses build upon the existing theoretical models that examine consumer substitution between online and offline channels. In particular, our paper is most closely related to research on multichannel retailing that utilizes theoretical models of spatially differentiated commodity markets derived from Salop's (1979) circular city model (Balasubramanian 1998; Jeffers and Nault 2005; Viswanathan 2005; Cheng and Nault 2007; Guo and Liu 2007) and those derived from Hotelling's (1929) linear city model (Pan, Ratchford, and Shankar 2002, 2004; Chun and Kim 2005; Liu, Gupta, and Zhang 2006; Moorthy and Zhang 2007). Common assumptions in all of these models are the presence of transportation costs when consumers use the offline channel and some disutility costs of buying online. In some cases the size of the transportation costs plays a key role in determining the equilibrium that prevails in these models. For example, one stream of this literature examines the decision of incumbent retailers and new firms to enter the online direct channel: this decision often turns on the magnitudes of transportation costs, online disutility costs, and consumer reservation values (Liu, Gupta, and Zhang 2006; Cheng and Nault 2007; Moorthy and Zhang 2007). Thus, our paper provides empirical evidence for a set of assumptions that play a key role in many theoretical models. High search costs (and the role of the Internet in reducing them) also play a key role in models that examine the role of lower search costs on buyer and seller behavior (Bakos 1997; Zettelmeyer 2000; Cachon, Terwiesch, and Xu 2007). For example, Bakos (1997) uses the circular city model to examine the extent of product differentiation in online markets driven by the existence of separate search costs for product information and price information. Implicit in his model is the notion that consumers incur search costs in finding product information in offline stores. These search costs can be viewed as another form of transportation costs and therefore they also help drive our hypotheses.

2.1. Hypotheses on relative benefit of buying online

As noted above, the core conceptual framework in our paper is derived from spatial models of competition that include a direct marketer, in particular Balasubramanian's (1998) circular city model of offline retailers with a direct retailer in the center. Balasubramanian's model includes several key assumptions that motivate our first hypothesis; in later hypotheses we relax or extend these assumptions. Consumers buy a single standardized product, and have complete information about product availability (the product is stocked at all retailers) and prices. Consumers face a finite cost of traveling to traditional retailers that depends on their distance to the retailer. Therefore there is heterogeneity across consumers in the cost of buying offline that depends on their location. These costs may include monetary costs of travel as well as inconvenience costs and the opportunity cost of time. Consumers have a high reservation price relative to their transportation cost (i.e., the market is covered or the product is "popular"). In contrast, all consumers face an identical fixed cost of buying from a "direct retailer" or online channel. The fixed cost may include a shipping cost, an inability to carefully assess product quality, and a lack of immediate gratification. Further, in contrast to Viswanathan (2005), there exist no switching costs or network externalities that increase the costs of switching channels. We state all of our hypotheses in terms of a single representative consumer.

Consumers maximize utility by choosing between the offline and online retailer based on prices, offline transportation costs, and online disutility costs. All else equal, reductions in transportation costs directly increase the utility of purchasing from the offline retailer, and should decrease the likelihood that the representative consumer buys from the online retailer. To our knowledge, this direct test of the role of distance in the Balasubramanian model has not been performed in any prior work.

Hypothesis 1a: ***Convenience for popular products.*** *As distance to offline stores decreases, the likelihood of purchasing a commodity product online decreases.*

We also examine the impact of distance to the offline retailer on products that are not stocked in all offline stores. We label such products "less popular." While not previously emphasized in the literature, product selection may be an important competitive decision variable between the online and offline channels. Hypothesis 1a is motivated by theory that assumes two key aspects of consumer behavior. First, consumers have a high reservation price relative to their transportation costs and product prices. In other words, the market is fully covered. Second, consumers are fully informed about the prices and availability of products in the online and offline channels. This setting is similar to the market for best-selling books. For less popular products, the reservation value for the representative consumer is lower. Moreover, consumers are less certain about the availability of the product at the offline retailer. This can be viewed as an increase in average offline transportation costs (in expectation) for a given product. Thus, for less popular products, markets may be uncovered. As Cheng and Nault (2007) note, an example of such a market might be that for ethnic books in the US. In such a setting, reduction in the distance to offline stores has a weaker effect on the likelihood that a representative consumer buys online for two reasons. First, the reservation value of the consumer is lower, so changes in transportation costs have a smaller impact on the likelihood of buying online. Second, the likelihood that any given store has the less popular product is smaller, so the expected transportation cost declines less than if the product was a popular one (and was certain to be available at the offline retailer).

Hypothesis 1b emphasizes that the online-offline tradeoff is only relevant when the offline store has a given product in stock. Further, the likelihood that a given product is in stock depends on the breadth of product selection at the store in addition to the popularity of the product.

<u>**Hypothesis 1b:**</u> ***Product selection.*** *A decrease in distance to offline stores has a smaller impact (less negative in magnitude) on online purchases when the product is less likely to be stocked offline.*

For example, take a book that is likely sold at a large specialty store such as Barnes & Noble but not at a discount store such as Wal-Mart. Hypothesis 1b implies that the effect of Barnes & Noble on online sales of this book is larger than the effect of Wal-Mart. It is a version of the convenience hypothesis 1a, but it takes into account the fact that not all kinds of stores stock all products.

Our next hypothesis examines the role of online and offline prices on channel choice. In Balasubramanian's model, changes in online price directly influence the utility of buying offline, and vice-versa. That is, there exists a significant cross-price elasticity across the online and offline channels. Prior work has tested for and found such a cross-price elasticity in computers (Goolsbee 2001; Prince 2007), DVDs (Chiou 2005), and computer memory (Ellison and Ellison 2006), so while we incorporate cross-price elasticity in our econometric model we do not include it as a separate hypothesis.

Instead, we focus on how distance to retail stores is associated with changes in consumers' sensitivity to price. Decreases in distance to offline stores will, as before, increase the utility of buying offline. This makes a given representative consumer less sensitive to changes in online price. Put another way, a marginal consumer who would have previously switched to the online channel after a fall in the online price no longer does so. Therefore, the impact of online discounts is tempered by the existence of local retail stores. To our knowledge, this hypothesis has also not been explored in prior theoretical or empirical literature.

<u>**Hypothesis 2:**</u> ***Price.*** *As distance to offline stores decreases, online price decreases have a smaller (less positive in magnitude) impact on the likelihood of purchasing a commodity product online.*

2.2 Effects of changes in three assumptions on main hypotheses

In this section we examine how variations in the basic assumptions about offline transportation costs, consumer reservation values, and online disutility costs in the Balasubramanian (1998) model influence our predictions regarding Hypothesis 1a and 1b.

First, we examine whether changes in the distance referred to in Hypothesis 1a influence the likelihood of buying online. This is a statement about the second derivative. Hypothesis 1a can be seen as a statement about the first derivative: how changes in distance influence the probability of buying online. Here, we wish to see how this derivative changes with initial distance. We expect the effects of decreases in store distance to be greater when initial distance is longer; that is, we expect transportation costs in this setting to be convex. This hypothesis is motivated by several streams of prior research. Bresnahan and Reiss (1991) show that the impact of the marginal entrant on local market competition is declining in the number of existing competitors. Campbell and Hopenhayn (2005) find that competition is tougher in larger markets.

Hypothesis 3: ***Changing effects of distance.*** *A decrease in distance to offline stores has a larger impact (more negative in magnitude) on online purchases when the initial distance to offline stores is longer*.

Hypothesis 3 focuses only on popular products because, as noted above, the market for less popular products may not be covered and because availability of less popular products may be correlated with distance. Hypothesis 3 examines how transportation costs differ across distances without being confounded by issues of availability.

As noted in the discussion prior to Hypothesis 1b, the relationship between retailer distance and the likelihood of buying a product online is weaker for less popular products than for popular products. One reason for this weaker relationship is that the representative consumer's reservation value for less popular books is lower on average. For example, if the product is very unpopular, in the limit the reservation value is close to zero so that decreases in retailer distance have no effect on the likelihood of purchase, even when the product is stocked in stores. In other words, when the reservation value is low enough, no one buys the product, online or offline. Thus, an increase in reservation value, or preference for less popular products, increases the strength of the relationship between store distance and likelihood of online purchase for less popular products. We are able to observe such variation in reservation prices because consumer preferences vary across markets. In small markets with homogeneous tastes, reservation values and sales for less popular products are likely to be relatively low, both online and offline.[1] The purpose of this hypothesis is to show that the value of an online retailer's product selection varies across locations.

Hypothesis 4: ***Demand heterogeneity and selection.*** *A decrease in distance to offline stores will have a larger impact (more negative in magnitude) on online purchases of less popular products in markets with heterogeneous tastes.*

Our final hypothesis relates to taxes. While hypothesis 3 and 4 examined changes to consumer offline transportation costs and reservation values, this hypothesis examines changes to the disutility of buying online. For online commerce, sales taxes are assessed only in the set of states in which the online retailer has a physical presence (Goolsbee 2000). In those states, the disutility of buying online is greater.

---

[1]Admittedly, the products available in local stores could be endogenous. Still, we think treating product line breadth as exogenous to location is a reasonable simplifying assumption in our analysis for three reasons. First, the variation in store size in our data is relatively low. For example, based on press releases from Borders books we found that the average store size in locations without a university is 20,875 square feet, while size for locations with a university is 21,369 square feet. Further, store size in counties with under 100,000 in population was 18,917 square feet, while that in counties with population over 1 million was 23,250 square feet. Second, the number of books stocked in the stores do not appear to change across locations in a way that will affect our results. The average large bookstore stocks between 40,000 and 100,000 books (Brynjolfsson, Hu and Smith 2003) and more specifically, the average Barnes and Noble stocks between 60,000 and 200,000 books. Approximately 50,000 of those books are common across Barnes and Noble stores in all locations (Rosenthal 2005). Since our analyses focus on books with national sales rank less than or equal to 15,000, there is likely to be substantial overlap between our sample and the set of books that are common to stores in all locations. For discount stores, press reports have listed the number of books to be "under 2000" (Kirkpatrick 2003) and between 1000 and 1500 (Wagner 2003). Further, we visited a Wal-Mart and a Target store in each of Atlanta and New Jersey. The Wal-Mart had 898 books in Atlanta and 860 books in Kearney, NJ. The Target had 1373 books in Atlanta and 1195 books in Jersey City. Third, we investigated the possibility that product selection may be idiosyncratic for book stores in university towns that also serve as university bookstores. We found that none of the entering book stores in our data are university bookstores. Nevertheless, to the extent that these examples do not address all endogeneity issues, our results on product line breadth across types of locations should be interpreted with caution.

Hence, we expect that estimates of parameters measuring the trade-off between transportation costs and disutility of buying online to be higher in states where online sales taxes are assessed. Intuitively, competition from offline stores is fiercer in sales tax states.

Prior work has demonstrated that the assessment of online sales taxes influences consumer propensity to buy online (Goolsbee 2000; Ellison and Ellison 2006). In this way, we confirm an important result in the existing literature. Still, it is not our primary goal to assess the impact of sales taxes on electronic commerce. Rather, we see changes in sales taxes as another comparative static on the assumptions of the basic Balasubramanian (1998) model, allowing us to observe a change in the disutility of buying online. The hypothesis argues that offline transportation costs matter more when the disutility of buying online increases.

Hypothesis 5: ***Taxes.*** *A decrease in distance to offline stores has a larger impact (more negative in magnitude) for consumers in states where online sales taxes are assessed.*

3. Data Description

To examine how online behavior varies with offline supply conditions, we require detailed data on how consumer online purchases vary across local geographic markets. The data we use are online book purchases from Amazon.com.

Why Books? The online book market is particularly well-suited to studying how the value of buying online depends upon where you live. First, many of the core assumptions of the basic Balasubramanian (1998) and other spatial competition models are met. In particular, books are commodity products wherein brand-specific or product-specific factors are less likely to influence consumer substitution across channels. Second, purchase-related attributes that cannot be determined digitally (Lal and Sarvary 1999) are relatively unimportant in the book market, enabling us to focus on location-related factors. That is, online disutility costs are low relative to other markets, so our estimates of the magnitude of online disutility costs relative to offline transportation costs are likely to be conservative, in comparison to other products and industries. Third, the market for digitized books remains relatively small at 0.1 % of total book industry sales in 2004 (Fishman 2005) and 0.3% of total book industry sales in 2005 (Carvajal 2006); if this were not so, then the relationship between distance and online book sales would be confounded by digital downloads. Fourth, because books are inexpensive commodity products, they are representative of a wide variety of other commodity products available online, including DVDs, CDs, groceries, office products, and others. Even in categories like furniture and clothing where the disutility costs of buying online may be much higher than in books, our main result documenting the role of location and transportation costs is likely to still hold, however consumer sensitivity to changes in offline store distance may be smaller. Fifth, books are one of the few product categories (besides travel services and computer hardware) where online sales reached over 10% of total retail sales by 2005 (US Census 2007). Sixth, Amazon.com is by far the largest online retailer of books with over 70% market share in the online world (Ehrens and Markus 2000; Weber 2005). Therefore, in books (but not other categories) it is reasonable to use a single retailer like Amazon to examine general trends in the online market (and to view it as the single direct marketer in the Balasubramanian model). And finally, the main offline retail stores with whom online stores compete are clear in the book market and we have precise data on when these stores open in a given location. Consequently, we can set up an effective natural experiment to explore channel substitution.

Purchase Circles Data: An observation in our data consists of a particular product-location-time. The data that we use come from the web pages on “Purchase Circles” from the Amazon.com web site. Amazon’s Purchase Circles are specialized best-seller lists that denote the top-selling books by location throughout the US. Henceforth, we use the word *locations* to refer to small and large cities, as well as small towns. We used a JAVA “spider” to extract and parse data from Amazon’s website. Between April 2005 and January 2006 this “spider” program visited Amazon’s website monthly and collected monthly data on purchases for each location in the Purchase Circles.[2]

[2] Some locations in our Purchase Circles raw data set do not appear for the entire time period. This matters if locations appear in Purchase Circles only when there are a sufficient number of purchases at Amazon because it is possible that local entry by retail stores may influence whether a location appears at all. In particular, due to a managerial decision at Amazon related to the threshold for inclusion in Purchase Circles, the number of locations in

While previous studies have used data from Amazon, our use of the Purchase Circles data to understand channel substitution is unique. To our knowledge the only other study to use the data available through Purchase Circles is Forman, Ghose, and Wiesenfeld (2006). However, they use the data to study the relationship between product reviews and sales. Another study using similar data is Bajari, Fox, and Ryan (2006) who use sales rankings of mobile phone carriers in 22 US markets to examine market power.

For each location, Amazon provides a list of the top 10 selling products. 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 products that appear in the top 10 in a location, there is considerable heterogeneity in this measure across locations and over time. Consumers buy different products in different locations. Figure 1 shows that in May 2005, 58.6% of products in our sample appear in the top 10 products at five or fewer locations, while only 1.5% of products appear more than 1000 times.

The use of rank data, rather than quantity data, means that our empirical framework is different than those typically used to examine channel substitution. In particular, it means our analysis must be based on relative rather than absolute sales. Therefore we translate our hypotheses into testable implications of how the relative sales of popular and less popular products vary across locations. These testable implications arise from the fact that while sales of popular and less popular products are sensitive to variations in local retail store distance, sales of unpopular products that are not stocked in local retail stores are not. Hypotheses on the likelihood that a particular book is purchased online are therefore translated into testable implications of the likelihood that a particular product appears in a local top 10. For example, for Hypothesis 1a we test whether the likelihood of a popular product appearing in the local top 10 declines as distance to an offline store decreases.

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

---

the sample expanded significantly in November 2005. For this reason, we only include locations that are observed before and after this date. This resulted in 1497 locations.

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 from Amazon on the 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. Shipping costs are identical across locations and are therefore not collected.

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 list price, normalized by the list price. The list price is the "recommended" price for a book that is typically printed on the book itself. Bestsellers are usually discounted, both online and offline, while older products typically sell at this list price. While our information on list price is taken from Amazon's website, the list price is publicly known and easily verified.[3]

In addition to price, we examine the national rank (popularity) of a book on Amazon. To allow for a flexible functional form, we compute a series of dummy variables (a spline) that indicate the specific range of national sales rank for which the book appears in that month: top 150, 151-500, 501-1500, 1501-5000, 5001-15,000, or greater than 15,000 (which we use as the base). We also estimate a log-linear specification. We define very popular products as those that fall in the top 150 nationally and popular products as those that fall in the range 151-500. Products with national sales ranks in the lower ranges, specifically those not in the top 1500, are classified as somewhat less popular (1501-5000) and less popular (5001-15,000) products. We supplement this with two alternative definitions of overall popularity: appearance on a New York Times Bestsellers list that month and appearance on a USA Today ranking of the top 150 books in the United States. We focus on the Amazon rankings because they provide detail on the rank of all products, not just the most popular.

To construct our final data set at a product-location-month level, for each month we identified the

[3] To verify the accuracy of our price data, we took a sample of 83 products from the NY Times bestseller lists (and that appeared in our data) and compared the Amazon list prices to a well known-book shopping bot, Findbookprices.com. 72 of the 83 books had list prices that were exactly the same, and the average deviation for prices that differed was just 6.3%. This may be due to the two year lag between our initial data collection and our price accuracy checks, and the fact that publishers change list prices from time to time (Borders 2007).

300 products that were most frequently listed in the local top 10 lists. We constructed an "outside option" of products that were listed in a local top 10 but were not in this group of 300. This outside option product had characteristics equal to the average of products in this set. We also provide a specification with the 1000 products that appeared most frequently in the top 10 lists. When deciding on the size of the set of products to include in the choice set, we had to make a tradeoff between two competing objectives. On the one hand, to identify whether product selection matters (Hypothesis 1b), we wanted to make the size of the choice set as large as possible. On the other hand, if we made the choice set too large, then we would have many products that are rarely in a local top 10. Since our product-location fixed effects rely on differencing dependent and independent variables from mean values, this too is unappealing. We chose 300 because we judged it the best compromise in this tradeoff.

Store Entry and Location-Level Data: Our main analysis examines how offline retail store entry influences buyer choice online. Retail store entry in a given location decreases the average distance consumers in that location must travel to access offline retailers, and also increases the availability of any given product, other things equal. 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 and zero otherwise; our variable labeled *Large Bookstore Entry* is equal to one for every month after a Barnes & Noble or Borders bookstore has entered within a 5.4 mile radius of the location and zero otherwise. These data were collected through press releases from the companies and through direct communication with company representatives. To compute radii, we use the average longitude and latitude across zip codes within the location. We use 5.4 miles because this is the distance that the average consumer travels to go to a bookstore (Brynjolfsson and Smith 2000), although we show that the results are robust (and in fact stronger) when we use a larger radius of 20 miles. Across our entire sample, 16.4% of locations experience discount store entry, while 4.7% experience a large bookstore entry. We focus on these particular stores because they represent the top two bookstores and the top two retailers who sell books. Based on reports from company financial reports, as well as media articles from Internet Retailer and BGI Media Center, we found that in 2006

Barnes & Noble (the largest book seller) had sales of $4.1 billion, Borders (the second largest book seller) had sales of approximately $3.4 billion, and sales of the third largest bookstore chain, Books-a-Million totaled approximately $470 million. Furthermore, Publishers Weekly (2006) emphasizes that Wal-Mart "can account for 40% of sales of bestselling books" with an overall book market share "as high as 10%". Unfortunately we could not find data on Target's sales of books. Still, Publisher's Weekly (2007) cites the abundance of Target stores in the Minneapolis area as one reason why small bookstores have struggled there.

In addition to the store entry data, we also collected location-level information on population using Census Bureau estimates for 2004 , on whether the location has a university from Barron's educational series, and on the number of local broadband providers in the location from semiannual Federal Communications Commission Form 477 data from December 2004, June 2005, and December 2005. Number of broadband providers is equal to the value of the last semiannual survey, though we also experimented with imputing the non-survey months through regression methods and linear interpolation.

4. Econometric Model

As discussed above, we examine the tradeoff between the transportation and search costs of buying offline and the various disutility costs of buying online. Identifying this tradeoff, however, is difficult to do in practice. In particular, it is difficult to separately identify supply and demand effects. For example, large cities may differ from small towns because (1) There are more stores in large cities (supply), or (2) People in large cities have different tastes than people in small towns (demand). One solution would be to directly measure the number of stores in each location and to regress sales rank on number of stores and several controls for demographics to attempt to control for taste. However, it is likely that this would suffer from the same difficulty: locations with more bookstores are likely those locations where many people buy books and therefore there are more bookstores because of local tastes. Consequently, separating out the effect of interest (how local competition affects online purchases) from other effects such as demand variation cannot be done in a simple cross-section.

Another solution common in the economics literature is to use an instrumental variables approach: if we could identify something that is correlated with the number of stores in a market but not with local demand then we could use that to identify the effect of the number of stores on online purchases. Unfortunately, we do not have access to such an instrument. Local characteristics associated with the number of stores selling books (e.g. population, income, and education) are generally correlated with local preferences for books.

For these reasons, we develop a panel data "difference-in-difference" approach. The panel data allow us to control for all location-specific preferences by including dummy variables for each book in each location. Our analysis then asks: how do purchases of product *j* in location *l* change after a new retailer enters location *l* compared to changes in purchases of product *j* in location *m*. Store entry can be viewed as a "natural experiment" where locations that experience entry are the "treatment group" and locations that do not experience entry are the "control group". Just as in an experimental setting, we examine whether the behavior of the treatment group changes in a way different from the behavior in the control group. Table 2 illustrates this logic. Locations that experience entry change their purchase behavior from C (before entry) to D (after entry). Over the same time period, locations that do not experience entry change their purchase behavior from A to B. The comparison between the treatment and control groups is then (D-C)-(B-A).

This difference-in-difference approach is commonly used in economics to examine the impact of location-level policy changes on outcomes. For example, Athey and Stern (2002) examine the effects of information technology adoption in 911 emergency response centers. Adoption of a new sophisticated 911-related information technology varied across Pennsylvania counties in the mid 1990s. They show that counties that adopted the technology (the treatment group) had larger improvements in medical outcomes than counties that did not adopt (the control group).[4] We use a similar intuition: locations that experience

[4] Another example is Milyo and Waldfogel (1999). They examine the impact of advertising on prices. They utilize a Supreme Court decision that lifted a ban on advertising alcohol prices in Rhode Island. Rhode Island received the "treatment" of a change in alcohol price advertising. Massachusetts is their control group. They show that the change in alcohol prices in Rhode Island was distinct from that in Massachusetts.

entry are considered the treatment group while other locations are the control group. Our interpretation of entry is that it represents a reduction in the average distance for local individuals to travel to a retail store that sells books.

In particular, we estimate the following linear probability model of whether a product *i* is in the top 10 in location *j* in month *t*:

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

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* in month *t* or earlier; $\boldsymbol{NationalRank_{it}}$ is a vector of dummy variables for the national sales rank of product *i* in month *t* defined above; $RelativePrice_{it}$ is the online price relative to the list price as defined above; X_{it} are other attributes of product *i* for month *t*;[5] μ_{ij} is a product-location fixed effect, μ_t is a month fixed effect, and ε_{ijt} is a product-location-month idiosyncratic error term. The product-location fixed effect, μ_{ij}, controls for the overall preferences of each location for each product. The entry interactions generate the natural experiment. We estimate this regression by differencing the average values across product-location. This method means that the calculated "within" R-squared values do not take into account the explanatory power of the fixed effects. Therefore, for our main results, we also estimated an equivalent, though computationally inefficient, "one-way fixed effects" estimator in order to calculate R-squared values that include the fixed effects.

The key assumption in difference-in-difference estimation is that unmeasured factors affect the

[5] These include ratings and number of reviews on Amazon, the number of days since the book launch, and a measure of local broadband competition. Also, the price information is missing for a few products. For these cases, we include a dummy variable indicating a "missing price" in order to reduce any potential impact of the missing observations on the price coefficients. In the Appendix we include a number of specifications to examine the robustness of our results to different ways of treating missing prices, including case-wise deleted and imputation. The results remain qualitatively similar.

treatment and control groups equally. While the product-location fixed effects in our model control for possible differences between the treatment locations (that experience entry) and the control locations (that do not), if areas that experience entry are also experiencing a change in local demand preferences then the treatment group changes over time in a different way than the control group. While we provide some evidence that this is not driving our results in the robustness section by showing that the results hold in both high growth and low growth locations, this is a necessary maintained assumption in interpreting our results.

There are two additional properties of our empirical framework that are important to discuss before we present our results. First, our coefficients of interest are on interaction terms. This means that using a non-linear model such as a Probit would be difficult to interpret because the cross-partial of a non-linear model may have a different sign than the coefficient on the interaction term (Ai and Norton 2003). The main disadvantage of using a linear model is reduced efficiency. Given the large number of observations in our study, this is less important. Second, as suggested by Bertrand, Duflo, and Mullainathan (2003), our difference-in-difference estimates may overstate the significance of the results without a standard error correction that addresses the fact that a given location is counted several times (i.e. for many products) in the data even though entry occurs just once. For this reason, we use White's heteroskedasticity-robust standard errors and cluster by location-month. We also experimented with clustering over product-months and the results are qualitatively similar.[6]

Our hypotheses from Section 2 easily convert into testable hypotheses on the coefficients on the interaction of local supply characteristics and product characteristics. Table 3 summarizes these coefficients and our results. Hypothesis 1a suggests that decreases in distance to offline stores are associated with relatively fewer purchases of popular products online. Entry by any type of store decreases such distances, other things equal. Therefore the coefficients on the interactions of *DiscountStoreEntry* or *LargeStoreEntry* with our *NationalRank* dummies for products that are nationally

[6] The robust standard errors also address the possibility 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.

in the top 150 and in the 151-500 range are hypothesized to be negative. Hypothesis 1b looks at product selection. Since large bookstores have a larger selection than discount stores, we expect large bookstore entry to have a larger impact on the less popular (i.e. nationally ranked in the 5000 to 15000 range) and somewhat less popular products (in the 1500 to 5000 range) than discount store entry. This range is chosen because while the typical Wal-Mart has just 1000-5000 books, the typical Barnes & Noble or Borders has a much higher inventory of books (Brynjolfsson, Hu, and Smith 2003).[7] Therefore, we expect the coefficient on the interaction of *LargeStoreEntry* with *NationalRank* products in the 5000 to 15000 range to be more negative than the coefficient on the interaction of *DiscountStoreEntry* and the products in this range. Hypothesis 2 suggests that entry by discount stores and large stores mitigates the effect of online price discounts because they discount the same types of books as the online retailer: i.e. the interactions of *DiscountStoreEntry* or *LargeStoreEntry* with *RelativePrice*. Hypotheses 3 through 5 are identified from differences in the magnitude of these coefficient estimates for different locations: small markets versus large markets; university towns versus those without a university; and locations with online sales taxes versus those without.

5. Results

5.1 The relative benefit of buying online

In this section, we show that changes in distance to local retail stores have a substantial effect on the types of products that appear in a local top 10 list. As discussed in sections 2 and 4, hypothesis 1a implies that, because of improved convenience, entry by discount stores and large bookstores should decrease purchases of popular products that are more easily available offline. Figure 2 provides suggestive evidence that the popularity of products appearing in a local top 10 list declines substantially after a store enters a local market. In particular, for those locations that experienced entry, the fraction of products that appear in a local top 10 that are also part of Amazon's nationwide top 1500 declines from 16% prior to Walmart or Target entry to 12% post-entry. Thus suggests that local store entry decreases the purchases of popular commodity products that are more easily available offline.

[7] See footnote 1 for further detail.

Our main regression results in Table 4 column 1, where we control for location-specific product preferences, are consistent with the descriptive statistics. Rows 1 and 8 of Table 4 show this most strongly: discount store and large bookstore entry decrease the likelihood of a local top 10 appearance by products in the national top 150 by 3.2 and 3.4 percentage points respectively. These results are significant at the 1% level and economically large relative to the average likelihood that a national top 150 product appears in a local top 10 (9.8%). This suggests that online disutility costs are substantial, and changes in the distance to offline stores appear to shape consumers' channel choice.

Table 4 column 1 provides little evidence that changes in retailer distance affect the decisions of consumers to purchase less popular and somewhat less popular products. Hypothesis 1b implies that the marginal effect of store entry over the range of these products is greater for big bookstores than for discount stores. Our test of the selection effect relies on the examination of the difference between discount store and large bookstore entry. In particular, we argue that selection would imply the entry interaction coefficient on less popular products (i.e. with national popularity of 5001-15,000) and somewhat less popular products (i.e. with national popularity of 1501-5000) should be more negative for large bookstores than for discount stores. As discussed above, we focus on these products because they are likely to be stocked in large bookstores but not in discount stores. In Table 4, we do not find evidence consistent with the selection hypothesis: the coefficients in rows 5 and 12 (or rows 4 and 11) are not significantly different from each other.

Figure 3 graphs the marginal effects of these interaction coefficients relative the base of products not in the national top 15000. It provides a visual representation of the results in Table 4, and shows that most of the impact of new store entry is found among the most popular products in the national top 150.

We next examine how offline store entry influences the effectiveness of online price discounts. Before discussing this interaction, we note that the negative sign in row 15 confirms the cross-price elasticity results of prior literature (e.g., Goolsbee 2001) - price discounts increase relative sales. Hypothesis 2 conjectures that as distance to offline stores falls (and hence the transportation cost associated with buying offline falls), online discounts become less effective. Rows 6 and 13 of Table 4 show 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. In the absence of retailer entry an Amazon discount relative to list price has a coefficient of -0.0237 (row 15). In contrast, when a discount store enters, this effect reduces to -0.0090 (row 6 plus row 15) and when a large bookstore enters it reduces to -0.0054 (row 13 plus row 15). So, price discounts have less of an impact on the likelihood of a product appearing in a local top 10 in the presence of store entry. This provides evidence in support of Hypothesis 2: as the distance to offline stores falls, consumers become less sensitive to online discounts.

These results are robust to a variety of different specifications. In column 2 of Table 4 we show the results of changing the entry radius from 5.4 to 20 miles. This threshold was based on the findings of Brynjolfsson and Smith (2000) who find that 8% of consumers live more than 20 miles away from the nearest general selection bookstore. The results are qualitatively similar to those in column 1. In column 3 we examine the use of a continuous measure of book popularity, log (Sales Rank), rather than a spline. While this measure does not allow us to examine Hypothesis 1b on selection, the convenience and price results are qualitatively unchanged. In column 4 we examine the robustness of our results to using a choice set of 1000 products rather than 300. Again, the results are qualitatively similar to column 1.

Our results are also robust to numerous other specifications (shown in the appendix), including different distance measures, different definitions of the timing of entry, a different definition of broadband diffusion, different location growth rates, different ways of treating missing prices, and different ways to define popular products including USA Today’s bestsellers list and the New York Times bestsellers list.

Overall, these results suggest that transportation costs impact online behavior. Improvement in the convenience of offline retail options is associated with a shift away from buying popular products online. New store entry also is related to reduced sensitivity to changes in online price. Since Amazon discounts best-selling products most heavily, this means that new store entry is associated with a shift away from popular products due to *both* convenience and price effects. Our results suggest, for commodity products that are ordered online and then mailed to consumers, a reduction in the offline cost of traveling to a retailer changes the products bought online and the impact of online price discounts.

5.2 Market characteristics and the relative benefit of buying online

Table 5 examines our remaining hypotheses. It shows how our results change with the size of transportation costs, local consumer preferences (or reservation values), and the disutility of buying online (as proxied by changes in sales tax rates). In Hypothesis 3, we argued that store entry should decrease purchases of popular products by more in small markets than in large markets. Our results for entry by discount stores (in columns 1 and 2 rows 1 and 2 of Table 5) support this hypothesis. Entry by discount stores decreases the likelihood of products in the national top 150 appearing in a local top 10 by 4.1 percentage points in small locations, compared to a decline of 3.0 percentage points in large markets. However, entry by large bookstores has little effect on consumer behavior in small markets, even for the most popular products. While we believe this is because there is little entry by large bookstores in these locations (and consequently the test has little power), we note that the results on large bookstores do not support the hypothesis.

Hypothesis 4 argues that markets with heterogeneous tastes are more likely to show a selection effect. The university location results in column 4 rows 5 and 11 show evidence supporting this. For university locations, coefficients on the interactions of less popular products (national rank 5001-15000) with large bookstore entry are more negative than for the interactions with discount store entry (with 5% significance in a Wald test). In contrast, in locations without a university, neither discount store entry nor large bookstore entry has any discernible impact on the purchases of lower ranked products. The somewhat less popular product results in rows 4 and 10 also show that the coefficients on the interactions are more negative in university locations than in those that do not have a university, though the university results are not significant in a Wald test. We also note that consumers in cities over 1 million (column 2) are more likely to display a selection effect than those in small towns (column 1). It appears that consumers use online retailers to achieve better selection in university towns and in large cities (where reservation values for less popular products may be relatively high), while consumers in small towns and towns without a university do not.

In Hypothesis 5, we argued that our results on channel substitution would be stronger in locations in which sales taxes are levied because the disutility of the online channel is higher. The results in columns 5 and 6 of Table 5 strongly support this assertion. Rows 1, 2, 7, and 8 show that the interaction of our popularity dummies with our entry variables is larger (in absolute value) for locations with online sales taxes than for states without such sales taxes. In fact, evidence of channel substitution is much stronger for products across all popularity ranks. As an example, row 7 shows that entry by large bookstores decreases the likelihood that products in the national top 150 would appear in the local top 10 by 10.0 percentage points in sales tax locations (1% significance), while only decreasing the likelihood of a local top 10 appearance by 3.2 percentage points in other locations (1% significance). In sum, we find evidence that the effects of convenience are significantly stronger in sales tax locations.

6. Discussion

6.1 Implications for research

Our results provide empirical support for the assumptions of a widely used theoretical modeling framework: spatial differentiation models that include a direct channel (e.g. Balasubramanian 1998). We find that variables and parameters in these models such as offline transportation cost, online shopping disutility cost, market coverage, and the prices of online and offline retailers interact to determine consumers channel choice in a way that is consistent with these models. Moreover, our results are suggestive about the relative magnitudes of some of these parameters, showing that online disutility costs can be significant, even for products such as books for which non-digital attributes are relatively unimportant. Knowledge of the relative magnitudes of these parameters is important for determining the relative profitability of online and offline retailers (Balasubramanian 1998) and for also determining the attractiveness of entry into the online market for incumbent offline retailers and new entrants (e.g., Liu, Gupta, and Zhang 2006; Cheng and Nault 2007). This theory research on whether to operate in online or offline channels is important to retailers, as evidenced by the recent decision of Borders Group to close many offline stores and open its own a retail web site (Bosman 2007).

Our empirical results also identified a set of potentially useful extensions to these models. In particular, our results suggest the usefulness of (i) understanding when the substantially wider product availability in online stores can act as a potential deterrent for entry by offline stores or for a brick & mortar firm to establish a direct channel, (ii) incorporating the effect of varying offline transportation costs in making optimal product assortment decisions for commodity products in local as well as online stores, and (iii) incorporating the effect of product popularity in modeling the impact of product returns on retailers' pricing decisions since the cost of returns to retailers and to consumers are likely to vary by product popularity and distance to stores, respectively.

Our discussion of the roles of convenience and product selection for online buyer behavior also advances a small empirical literature on online-offline channel substitution that has thus far focused primarily on the ability of the online channel to offer lower prices (e.g., Prince 2007). Our results show that the convenience of the offline channel appears strongly related to online choices for very popular products; however, we also find the impact of offline channel on online purchases of less popular and somewhat less popular products is limited to large markets and university towns where tastes may be more heterogeneous.

6.2 Implications for managers

These results can have important managerial implications for online and offline retailers. For online retailers, our research shows how consumers' use of the online channel varies across locations. If consumers use internet channels primarily to obtain lower prices for or more convenient access to very popular commodity products, then the expansion 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. The presence of significant online disutility costs suggests that there is likely to be an upper bound on the extent to which consumers migrate from offline world to the online world to purchase commodity products. Our results can also inform estimates of potential revenues from electronic commerce in different locations. This can be especially useful for "brick & click" retailers who practice dual-channel marketing strategies, and thus need precise optimization strategies to prevent inter-channel cannibalization and maximize profits. For offline retailers, our work shows that online retailers are relevant competitors. Competition depends on more than the number of local stores, it also depends on product overlap and disutility costs associated with the online channel. This statement has direct practical relevance to practitioners and policy-makers: in 2005, the year of our data, the Federal Trade Commission blocked the proposed merger between Blockbuster and Hollywood Video partially on the assertion that competition from the Internet was irrelevant and only the number of local retailers mattered.

6.3 Limitations of Research

As with any empirical work, the data that we bring to bear has some limitations. For one, 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 less popular goods, we are limited in our ability to make inferences about purchases of very unpopular products. Similarly, while our sample is roughly representative of the US population in towns over 5,000 people—for example, the median of per capita income among such towns in our sample is $23,857 compared to $19,792 for the entire country, according to 2000 US Census place data—we can say little about the 11.6% of the population that lives in even smaller places. Thus we do not measure channel substitution in truly rural locations.

Also, we examine online behavior for only one particular product: books. As discussed earlier, we focus on books for a number of reasons: books are commodities whose characteristics do not vary by channel, books are representative of products that represent a substantial portion of total online retail sales, books are an area where online retailing is relatively important, Amazon is sufficiently dominant in online book sales that it is reasonable to use Amazon-only data to study books, and the key offline competitors are well-defined. Our results are likely to be particularly informative about other products that share similar characteristics, such as toys, many health and beauty products, and some electronics where the set of attributes is small and well-defined. However, key online retail categories such as travel, financial services, and automotive are sufficiently different from books that our results should be applied carefully to these settings. Examining other products and contrasting these with the book results may enable separation of location-specific effects from the product attribute effects discussed in Lal and Sarvary (1999).

Last, we are unable to observe the consumer-level decision to use the online channel. Our inferences are based on changes in the popularity of the products consumers purchase across locations and over time, but we have no information on the binary decision of a given consumer to use the online channel. Furthermore, we do not survey people about their purchasing motivations. In this way, our data do not allow us to observe the exact mechanism that drives the substitution. Our results are therefore only as strong as the arguments for revealed preference methods rather than survey-based methods (for one discussion of this see Manski 2000); in other words, our results depend on the degree to which people's actual behavior allows us to understand their underlying motivations and preferences.

6.4. Conclusion

Utilizing a unique panel data that illustrates differences in book-buying behavior across 1497 locations in the US, we provide evidence that distance to local retail stores shapes the way that consumers use the online channel. In particular, controlling for consumer preferences, we examine whether consumers with few local retail options purchase systematically more popular or less popular and more expensive or less expensive products than those with more local retailers. More generally, our research provides evidence of how changes in offline transportation costs and online disutility costs shape consumer decisions online. We draw on and inform a substantial literature on channel substitution to show that the benefit of buying online depends on where you live.

References

- Ai, C., E. Norton. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters.* 80, 123-129.
- Athey, S., S. Stern. 2002. The Impact of Information Technology on Emergency Health care Outcomes. *RAND Journal of Economics* 33, 399–432.
- Bajari, P., J.T. Fox, S. Ryan. 2006. Evaluating Wireless Carrier Consolidation Using Semiparametric Demand Estimation. NBER Working Paper # 12425.

- Bakos, J.Y. 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43, 1676-1692.
- Balasubramanian, S. 1998. Mail Versus Mall: A Strategic Analysis of Competition Between Direct Marketers And Conventional Retailers. *Marketing Science* 17, 181-195.
- Balasubramanian, S., P. Konana, N. Menon. 2003. Customer Satisfaction in Virtual Environments: A Study of Online Investing. *Management Science*. 49, 871-889.
- Baye, M., J. Morgan, P. Scholten. 2007. Information, Search, and Price Dispersion," in *Handbook of Economics and Information Systems* (T. Hendershott, ed.), Elsevier Press, forthcoming.
- Bitran, G., S. Mondschein. 1996. Mailing Decisions in the Catalog Sales Industry, *Management Science* 42, 1364-1381.
- Borders 2007. Searching for Books, Music and Movies. Available at http://www.bordersstores.com/care/care.jsp?page=2
- Bosman, J. 2007. Borders to Close Many Stores and Start a Retail Web Site. *The New York Times* (March 23), C2.
- Bresnahan, T., P. Reiss. 1991. Entry and Competition in Concentrated Markets. *Journal of Political Economy* 99, 977-1009.
- Bronnenberg, B., V. Mahajan. 2001. Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables. *Marketing Science* 20, 284-299.
- Brynjolfsson, E., M. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46, 563-585.
- Brynjolfsson, E., Y. Hu, M. Rahman. 2007. Battle of the Retail Channels: How Product Selection and Geography Drive Cross-channel Competition. Working Paper, Purdue University.
- Brynjolfsson, E., Y. Hu, M. Smith. 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety. *Management Science* 49, 1580-1596.
- Cairncross, F., 1997. The Death of Distance. Cambridge, MA: Harvard University Press.
- Cachon, G., C. Terwiesch, Y. Xu. 2006. On The Effects Of Consumer Search And Firm Entry On Multiproduct Competition, forthcoming, *Marketing Science*
- Campbell, J.R., H. Hopenhayn. 2005. Market Size Matters. *Journal of Industrial Economics* 53, 1-25.
- Carvajal , D. 2006. Springer Shifts Model For Sales of E- Books. *International Herald Tribune*, July 23.
- Cheng, Z., B. R. Nault. 2007. Internet Channel Entry: Retail Coverage And Entry Cost Advantage. *Information Technology and Management* 8, 111-132.
- Chiou, L. 2005. Empirical Analysis of Retail Competition: Spatial Differentiation at Wal-Mart, Amazon.com, and their Competitors. Working Paper, Occidental College.

- Chun, S.-H., J. Kim. 2005. Pricing Strategies in B2C Electronic Commerce: Analytical And Empirical Approaches. *Decision Support Systems* 40, 375-388

- Dellarocas, C. 2003. The Digitization of Word Of Mouth: Promise And Challenges Of Online Feedback Mechanisms. *Management Science* 49, 1407-1424.

- DSN Retailing Today. 2002. Top Book Retailers. (July 8).

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

- Ehrens, S., A. Markus. 2000. Amazon.com: There's an "R" in e-tailing. *Epoch Partners Consumer Internet Company Report* (November 13) 4.

- Fishman, R. 2005. Signs of Life for E-books in 2004. *Publishers Weekly*, (Jan. 17).

- Forman, C., A. Goldfarb, S. Greenstein. 2005. How did location affect adoption of the commercial Internet: Global village Vs. urban leadership. *Journal of Urban Economics* 58, 389-420.

- Forman. C, A. Ghose B. Wiesenfeld. 2006. Examining the Relationship Between Reviews and Sales: The Role of Social Information in Electronic Markets. NYU CeDER Working Paper # 06-09.

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

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

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

- Goolsbee, A. 2001. Competition in the Computer Industry: Online Versus Retail. *Journal of Industrial Economics* 49, 487-499.

- Guo, L., Y. Liu. 2007. To Restrain or to Expand: Optimal Retail Store Opening Strategies in Coping with Manufacturer Direct Entry. Working Paper, University of Illinois at Urbana-Champaign.

- Hotelling, H. 1929. Stability in competition, *Economic Journal* 39, 41-57.

- Jank, W., P.K. Kannan. 2005. Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice. *Marketing Science* 24, 623--634.

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

- Kirkpatrick, D. 2003. Shaping Cultural Tastes at Big Retail Chains. *The New York Times.* May 18.

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

- Liu, Y., S. Gupta, Z. Zhang. 2006. Note on Self-Restraint as an Online Entry-Deterrence Strategy. *Management Science* 52, 1799-1809.

- Manski, C. 2000. Economic Analysis of Social Interactions. *Journal of Economic Perspectives* 14, 115-136.

- Milyo, J., J. Waldfogel. 1999. The Effect of Advertising on Prices: Evidence in the Wake of 44 Liquormart. *American Economic Review* 89, 1081-1096.

- Moorthy, S., X. Zhang. 2007. Multichannel Retailing: Bricks-and-Mortar, Clicks, Or Bricks-and-Clicks? Working paper. University of Toronto.

- Pan, X., B. Ratchford, V. Shankar. 2002. Price Competition Between Pure Play Versus Bricks-And-Clicks E-Retailers: Analytical Model And Empirical Analysis. In *The Economics of the Internet and E-Commerce* (Advances in Applied Microeconomics, Volume 11), ed. M. Baye, Amsterdam, Elsevier Science, 29-62.

- Pan, X., B. Ratchford, V. Shankar. 2004. A Model of Retail Competition in Service and Price: Pure Play Internet vs. Bricks-and-Mortar Retailers. Available at http://ssrn.com/abstract=485902.

- Prince, J. 2007. The Beginning of Online/Retail Competition and its Origins: An Application to Personal Computers. *International Journal of Industrial Organization* 25, 139-156.

- *Publishers Weekly*. 2006. A Little Bit of Country. 243, August 21.

- *Publishers Weekly*. 2007. United States of Bookselling: Minnesota. 254, June 4.

- Rosenthal, M. 2005. Barnes and Noble Book Store vs. Amazon Stocking, www.fonerbooks.com, August 29.

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

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

- US Census. 2007. E-Stats Report. May. www.census.gov/estats

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

- Wagner, M. 2003. Where's Wal-Mart? *Internet Retailer*. On September 1, 2007, available at http://www.internetretailer.com/internet/marketing-conference/70624-wheres-wal-mart.html .

- Wall Street Journal. 2005. Editorial: Blocking Blockbuster. March 30, A16.

- Weber, S. 2005. Amazon Sales Rank and its Impact on Used Book Sales, December 01, available at http://weberbooks.com/selling/2005/12/amazon-sales-rank-and-its-impact-on.html

- Zettelmeyer, F. 2000. Expanding on the Internet: Pricing and Communication Strategies When Firms Compete on Multiple Channels. *Journal of Marketing Research* 37, 292-308.

Table 1: Summary Statistics for Books

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
BY LOCATION-PRODUCT-MONTH					
*Dummy for Top 10 in lo*cation	4,051,254	0.0347	0.1831	0	1
Relative Price	4,051,254	-0.2654	0.1434	-0.6	0
Very Popular Products (rank 1-150)	4,051,254	0.1711	0.3766	0	1
Popular Products (rank 150-500)	4,051,254	0.1737	0.3789	0	1
Moderately Popular Products (rank 500-1500)	4,051,254	0.1538	0.3608	0	1
Somewhat Less Popular Products (rank 1500-5000)	4,051,254	0.1351	0.3418	0	1
Less Popular Products (rank 5000-15000)	4,051,254	0.1296	0.3358	0	1
Unpopular Products (rank over 15000)	4,051,254	0.2367	0.4251	0	1
Dummy for missing price information	4,051,254	0.0644	0.2454	0	1
Average rating	4,051,254	4.1098	0.5617	1.5	5
Log(days since launch)	4,051,254	6.5007	1.4946	0	9.8268
Broadband	4,051,254	11.4887	3.3362	0	24
Dummy for missing elapsed date information	4,051,254	0.0259	0.1588	0	1
Log(Number of reviews)	4,051,254	4.9545	1.4596	0.6931	8.6500
Discount Store Entry within 5.4 miles	4,051,254	0.0809	0.2727	0	1
Large Bookstore Entry within 5.4 miles	4,051,254	0.0166	0.1276	0	1
BY LOCATION					
Discount store openings in all locations	1497	0.1643	0.3707	0	1
Discount store openings in small locations	143	0.0979	0.2982	0	1
Discount store openings in large locations	412	0.2087	0.4069	0	1
Large bookstore openings in all locations	1497	0.0468	0.2112	0	1
Large bookstore openings in small locations	143	0.0210	0.1438	0	1
Large bookstore openings in large locations	412	0.0752	0.2641	0	1
Location has a university	1497	0.4449	0.4971	0	1
Location in state with online sales tax	1497	0.0839	0.2782	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: Treatment and Control Groups

		BEFORE ENTRY	AFTER ENTRY
ENTRY	**NO (CONTROL)**	**A**	**B**
	YES (TREATMENT)	**C**	**D**

Table 3: Main Hypotheses and Summary of Results

Hypothesis	*Relevant Coefficients*	*Prediction*	*Intuition*	*Supported?*	*Location*
1a Convenience	• Very popular products*Large bookstore entry • Very popular products*Discount store entry • Popular products*Large bookstore entry • Popular products*Discount store entry	Negative	With an increase in the number of stores, more popular products are bought offline	Supported	Table 4 Column 1 Rows 1, 2, 8, 9
1b Product Selection	• Less popular products*Large bookstore entry • Less popular products*Discount store entry • Somewhat less popular products*Large bookstore entry • Somewhat less popular products*Discount store entry	Negative, but less so than very popular and popular products. More negative for large bookstores than for discount stores	With an increase in the number of large bookstores, more unpopular products are bought offline.	Not Supported in full data set	Table 4 Column 1 Rows 4, 5, 11, 12
2 Price	• Relative price*Large bookstore entry • Relative price*Discount store entry	Positive	The impact of online discounts is tempered by local retail stores.	Supported	Table 4 Column 1 Rows 6, 12
3 Changing effects of distance	• Very popular products*Large bookstore entry • Very popular products*Discount store entry • Popular products*Large bookstore entry • Popular products*Discount store entry	Negative, more so in small markets	Convenience effects are stronger in small markets where distance to stores is larger.	Partially Supported	Table 5 Columns 1, 2 Rows 1, 2, 7, 8
4 Demand heterogeneity and selection	• Less popular products*Large bookstore entry • Less popular products*Discount store entry • Somewhat less popular products*Large bookstore entry • Somewhat less popular products*Discount store entry	Negative in university towns especially, more negative for large bookstores than for discount stores	Selection effects are stronger in university towns where consumer preferences are more diverse.	Supported	Table 5 Columns 3, 4 Rows 4, 5, 10, 11
5 Taxes	• Very popular products*Large bookstore entry • Very popular products*Discount store entry • Popular products*Large bookstore entry • Popular products*Discount store entry	Negative in locations with online sales tax especially.	Convenience effects are stronger when fit of the online channel is worse due to sales taxes.	Supported	Table 5 Columns 5, 6 Rows 1, 2, 7, 8

Table 4: Main Results—Difference in Difference on store entry

Row			(1) 5.4 Miles	(2) 20 Miles	(3) Sales Rank	(4) Large Choice Set
1	Discount Store Entry (interactions)	Very Popular Products# (Top 150 nationally)	-0.0320	-0.0372		-0.0343
			(0.0012)**	(0.0008)**		(0.0011)**
2		Popular Products (151-500 nationally)	-0.0034	-0.0061		-0.0018
			(0.0008)**	(0.0005)**		(0.0005)**
3		Moderately Popular Products (501-1500 nationally)	-0.0060	-0.0080		0.0018
			(0.0006)**	(0.0003)**		(0.0003)**
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0082	-0.0084		-0.0007
			(0.0009)**	(0.0004)**		(0.0003)+
5		Less Popular Products (5001-15000 nationally)	-0.0019	-0.0020		0.0006
			(0.0007)**	(0.0003)**		(0.0002)*
6		Relative Price	0.0147	0.0107	0.0153	0.0166
			(0.0022)**	(0.0010)**	(0.0022)**	(0.0012)**
7		Log(Sales Rank)			0.0034	
					(0.0002)**	
8	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0339	-0.0343		-0.0387
			(0.0025)**	(0.0011)**		(0.0023)**
9		Popular Products (151-500 nationally)	-0.0029	-0.0045		-0.0017
			(0.0020)	(0.0008)**		(0.0012)
10		Moderately Popular Products (501-1500 nationally)	-0.0022	-0.0047		0.0041
			(0.0016)	(0.0006)**		(0.0007)**
11		Somewhat Less Popular Products (1501-5000 nationally)	-0.0074	-0.0067		0.0005
			(0.0025)**	(0.0009)**		(0.0009)
12		Less Popular Products (5001-15000 nationally)	-0.0023	-0.0022		0.0009
			(0.0018)	(0.0006)**		(0.0005)+
13		Relative Price	0.0183	0.0145	0.0228	0.0207
			(0.0061)**	(0.0021)**	(0.0063)**	(0.0031)**
14		Log(Sales Rank)			0.0035	
					(0.0004)**	
15	Other	Relative price (not interacted)	-0.0237	-0.0268	-0.0146	-0.015
			(0.0007)**	(0.0008)**	(0.0007)**	(0.0004)**
		Observations	4,051,254	4,051,254	4,051,254	9,420,562
		Number of Fes	978,611	978,611	978,611	2,933,794
		Within Estimator R-squared	0.07	0.07	0.07	0.04
		Fixed Effects R-squared	0.81	0.81	0.81	0.81
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Log(Sales Rank) (column 3) -Book popularity spline (columns 1, 2 & 4)	-Broadband competition -Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out)			

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects.
#Base is *unpopular products* ranked 15000 and up.
+ significant at 10%; * significant at 5%; ** significant at 1%. For columns 3 and 4, we use entry in a 5.4 mile radius.

Table 5: Results split by location size, whether the location has a university and whether the location has a sales tax

Row			(1) Locations Under 100k	(2) Locations over 1 million	(3) Locations without a university	(4) Locations with a university	(5) No Sales Tax Online	(6) Sales Tax Online
1	Discount Store Entry (interactions)	Very Popular Products[#] (Top 150 nationally)	-0.0410	-0.0299	-0.0291	-0.0348	-0.0315	-0.0458
			(0.0045)**	(0.0023)**	(0.0018)**	(0.0016)**	(0.0012)**	(0.0061)**
2		Popular Products (151-500 nationally)	-0.0071	0.0015	-0.0012	-0.0050	-0.0033	-0.0073
			(0.0030)*	(0.0016)	(0.0013)	(0.0010)**	(0.0009)**	(0.0032)*
3		Moderately Popular Products (501-1500 nationally)	-0.0065	-0.0023	-0.0040	-0.0075	-0.0057	-0.0123
			(0.0028)*	(0.0013)+	(0.0010)**	(0.0008)**	(0.0006)**	(0.0022)**
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0072	-0.0088	-0.0074	-0.0087	-0.0079	-0.0133
			(0.0050)	(0.0019)**	(0.0015)**	(0.0011)**	(0.0009)**	(0.0036)**
5		Less Popular Products (5001-15000 nationally)	-0.0006	0.0014	-0.0013	-0.0023	-0.0018	-0.0031
			(0.0037)	(0.0016)	(0.0011)	(0.0009)**	(0.0007)**	(0.0031)
6		Relative Price	0.0215	0.0240	0.0136	0.0158	0.0152	0.0068
			(0.0102)*	(0.0046)**	(0.0035)**	(0.0028)**	(0.0022)**	(0.0080)
7	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0068	-0.0388	-0.0276	-0.0427	-0.0320	-0.1004
			(0.0177)	(0.0045)**	(0.0033)**	(0.0039)**	(0.0025)**	(0.0252)**
8		Popular Products (151-500 nationally)	-0.0025	-0.0073	0.0034	-0.0119	-0.0022	-0.0466
			(0.0142)	(0.0034)*	(0.0027)	(0.0027)**	(0.0019)	(0.0238)+
9		Moderately Popular Products (501-1500 nationally)	0.0112	-0.0061	0.0033	-0.0097	-0.0016	-0.0472
			(0.0115)	(0.0026)*	(0.0024)	(0.0021)**	(0.0016)	(0.0214)*
10		Somewhat Less Popular Products (1501-5000 nationally)	0.0172	-0.0105	-0.0021	-0.0141	-0.0069	-0.0564
			(0.0116)	(0.0040)**	(0.0034)	(0.0038)**	(0.0025)**	(0.0243)*
11		Less Popular Products (5001-15000 nationally)	0.0167	-0.0049	-0.0008	-0.0043	-0.0021	-0.0368
			(0.0157)	(0.0024)*	(0.0031)	(0.0019)*	(0.0018)	(0.0042)**
12		Relative Price	0.0160	0.0041	0.0226	0.0130	0.0157	0.0696
			(0.0196)	(0.0090)	(0.0089)*	(0.0078)+	(0.0061)*	(0.0267)**
13	Other	Relative price (not interacted)	-0.0221	-0.0276	-0.0219	-0.0259	-0.0234	-0.0319
			(0.0023)**	(0.0015)**	(0.0010)**	(0.0012)**	(0.0008)**	(0.0040)**
		Observations	386,551	1,108,643	2,242,917	1,808,337	3,893,917	157,337
		Number of FEs	93,393	268,775	542,164	436,447	940,853	37,758
		Within estimator R-squared	0.07	0.08	0.07	0.07	0.07	0.08

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 4.
+ significant at 10%; * significant at 5%; ** significant at 1%. Entry is considered for stores entering within a 5.4 mile distance.

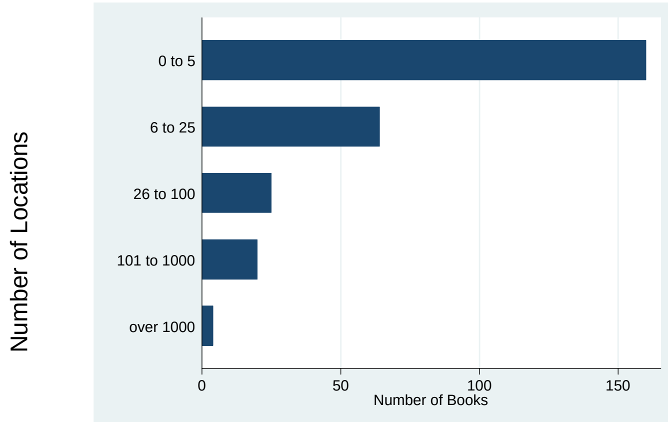

Figure 1: Number of Locations that a Product is in the Top 10--May 2005

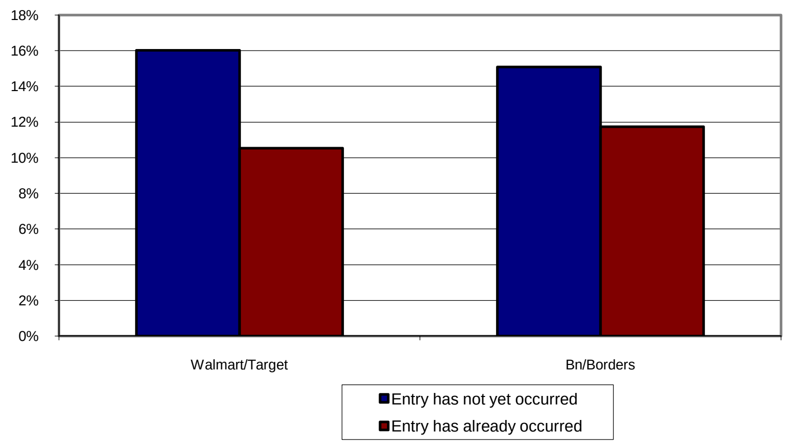

Figure 2: Likelihood of Product in Local Top 10 Appearing in National Top 1500, By Whether Entry Has Already Occurred by Store Type (Among Locations that Experience Entry, September)

Figure 3: Marginal Effects from Baseline Regression (Based on Table 4 column (1))

Appendix

This appendix includes additional checks of robustness to our core results. Appendix Table 1 shows that our results are robust to the use of different splines and measures of product popularity.[1] Appendix Table 2 shows that our results are robust to using absolute distance rather than the distance dummies (5.4 miles and 20 miles) that we use in our core results; in particular, stores that enter closer to the location will have a stronger effect on online behavior than those that enter farther away. Appendix Tables 3, 4, and 5 show that our results are robust to redefining entry as occurring one month after opening date, two months after opening date, and one month before opening date. These results demonstrate that noise in our measurement of the timing of entry would not influence our results. Moreover, they demonstrate that our results do not reflect simply a short-run effect of entry; the effect of entry remains even two months after store opening. Appendix Table 6 shows that our results are robust to different methods of estimating broadband penetration in months that do not coincide with the FCC's collection of Form 477 data. Appendix Table 7 shows that our results are robust to both high population growth and low population growth locations, and are not capturing a transition of an area being small and rural to becoming larger and more urban. Appendix 8 shows that our results are robust to different methodologies for treating missing observations.

[1] Note that there is substantial overlap in these different measures of popularity. For example, 95% of the books on the New York Times bestsellers list in November 2005 (and 82% of the books on USA Today's list) were in Amazon's national top 150.

Appendix Table 1: Difference in Difference on store entry: additional splines

Row			Spline 1#	Spline 2#	Spline 3#	USA Today#	NY Times#
			(1)	(2)	(3)	(4)	(5)
1	Discount Store Entry (interactions)	Very Popular Products	-0.0375	-0.0330	-0.0233	-0.0487	-0.0575
			(0.0014)**	(0.0013)**	(0.0010)**	(0.0018)**	(0.0022)**
2		Popular Products	-0.0060	-0.0044	-0.0067		
			(0.0008)**	(0.0008)**	(0.0007)**		
3		Moderately Popular Products	-0.0053	-0.0069			
			(0.0006)**	(0.0006)**			
4		Less Popular Products	-0.0063	-0.0091			
			(0.0006)**	(0.0009)**			
5		Even Less Popular Products		-0.0030	-0.0017		
				(0.0007)**	(0.0007)*		
6		Somewhat Unpopular Products		-0.0020			
				(0.0006)**			
7		Relative Price	0.0124	0.0141	0.0159	0.0060	0.0069
			(0.0021)**	(0.0022)**	(0.0022)**	(0.0021)**	(0.0021)**
8	Large Bookstore Entry (interactions)	Very Popular Products	-0.0357	-0.0351	-0.0238	-0.0597	-0.0505
			(0.0027)**	(0.0026)**	(0.0024)**	(0.0039)**	(0.0047)**
9		Popular Products	-0.0058	-0.0040	-0.0056		
			(0.0020)	(0.0020)*	(0.0016)**		
10		Moderately Popular Products	-0.0022	-0.0033			
			(0.0016)	(0.0016)*			
11		Less Popular Products	-0.0036	-0.0086			
			(0.0018)*	(0.0026)**			
12		Even Less Popular Products		-0.0036	-0.0020		
				(0.0020)+	(0.0017)		
13		Somewhat Unpopular Products		-0.0022			
				(0.0017)			
14		Relative Price	0.0170	0.0176	0.0201	0.0062	0.0135
			(0.0061)**	(0.0061)**	(0.0061)**	(0.0060)	(0.0059)**
15	Other	Relative price	-0.0165	-0.0237	-0.0254	-0.0214	-0.0170
			(0.0007)**	(0.0008)**	(0.0008)**	(0.0007)**	(0.0007)**
		Observations	4,051,254	4,051,254	4,051,254	4,062,326	4,052,722
		Number of FEs	978,611	978,611	978,611	981,255	978,611
		Within estimator R-squared	0.06	0.05	0.06	0.06	0.06
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Log(Number of reviews) -Broadband competition	 -Dummy for missing elapsed date information -Discount Store Entry within 5.4 (or 20) miles -Large Bookstore Entry within 5.4 (or 20) miles -Product-location fixed effects (differenced out) -Book popularity spline (cols 1 & 2 only) -Log(Sales Rank) (Column 3 only)				

#Spline in column (1) is top 100, 101-500,501-1000, and 1001-10,000; in column (2) is top 150, 151-500, 501-1500, 1501-5000, 5001-15,000, and 15,001-50,000; in column (3) is top 250, 251-5000, and 5001-15,000; in column (4) is top 150 books as according to the USA Today list of bestsellers; in column (5) is appearing in a New York Times bestsellers list. 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 2: Absolute Distance Results

Row			(1) 20 Miles	(2) Sales Rank
1	Discount Store Entry & Distance to Store (double interactions)	Very Popular Products# (Top 150 nationally)	-0.0009 (0.0001)**	
2		Popular Products (151-500 nationally)	0.0001 (0.0001)	
3		Moderately Popular Products (501-1500 nationally)	0.00002 (0.00005)	
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0002 (0.0001)*	
5		Less Popular Products (5001-15000 nationally)	-0.00002 (0.00005)	
6		Relative Price	0.0007 (0.0002)**	0.0009 (0.0002)**
7		Log(Sales Rank)		0.0001 (0.0000)**
8		Main Interaction: Entry * Distance to Store	0.0003 (0.0001)**	-0.0003 (0.0002)+
9	Large Bookstore Store Entry & Distance to Store (double interactions)	Very Popular Products (Top 150 nationally)	-0.0009 (0.0002)**	
10		Popular Products (151-500 nationally)	0.00002 (0.0001)	
11		Moderately Popular Products (501-1500 nationally)	0.0002 (0.0001)+	
12		Somewhat Less Popular Products (1501-5000 nationally)	-0.00007 (0.0002)	
13		Less Popular Products (5001-15000 nationally)	0.0001 (0.0001)	
14		Relative Price	0.0004 (-0.0004)	0.0007 (0.0004)
15		Log(Sales Rank)		0.0001 (0.0000)**
16		Main Interaction: Entry * Distance to Store	0.0002 (-0.0001)	-0.0005 (0.0003)+
17	Discount Store Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.03 (0.0012)**	
18		Popular Products (151-500 nationally)	-0.0073 (0.0008)**	
19		Moderately Popular Products (501-1500 nationally)	-0.0082 (0.0005)**	
20		Somewhat Less Popular Products (1501-5000 nationally)	-0.007 (0.0007)**	
21		Less Popular Products (5001-15000 nationally)	-0.002 (0.0005)**	
22		Relative Price	0.0051 (0.0017)**	0.001 (-0.0018)
23		Log(Sales Rank)		0.0037 (0.0002)**
24	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0278 (0.0018)**	
25		Popular Products (151-500 nationally)	-0.0047 (0.0013)**	
26		Moderately Popular Products	-0.0064	

		(501-1500 nationally)	(0.0009)**	
27		Somewhat Less Popular Products	-0.0067	
		(1501-5000 nationally)	(0.0014)**	
28		Less Popular Products	-0.0029	
		(5001-15000 nationally)	(0.0011)**	
29		Relative Price	0.0119	0.011
			(0.0033)**	(0.0035)**
30		Log(Sales Rank)		0.0033
				(0.0002)**
31	Other	Relative price	-0.0268	-0.0179
			(0.0008)**	(0.0008)**
		Observations	4,051,254	4,051,254
		Number of FEs	978,611	978,611
		Within estimator R-squared	0.07	0.07
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Book popularity spline (column 1) -Log(Sales Rank) (column 2) -Broadband competition	-Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out)	

Absolute distance is measured as the great circle distance using the latitude and longitude of the store and that of the location under observation. When then transform this to 20 – absolute distance (or in column 2, 5.4 – absolute distance) so that bigger numbers correspond to shorter entry distances. To compute radii, we use the average longitude and latitude across zip codes within the location. 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%. For sales rank regression, we use entry in a 5.4 miles radius.

Appendix Table 3: Alternate Entry (Only with a One Month Lag) Results

Row			(1) 5.4 Miles	(2) 20 Miles	(3) Sales Rank
1	Discount Store Entry (interactions)	Very Popular Products[#] (Top 150 nationally)	-0.0272	-0.0335	
			(0.0012)**	(0.0007)**	
2		Popular Products (151-500 nationally)	-0.0003	-0.004	
			(-0.0007)	(0.0004)**	
3		Moderately Popular Products (501-1500 nationally)	-0.004	-0.0069	
			(0.0005)**	(0.0003)**	
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0049	-0.007	
			(0.0008)**	(0.0004)**	
5		Less Popular Products (5001-15000 nationally)	-0.0019	-0.0016	
			(0.0007)*	(0.0003)**	
6		Relative Price	0.0036	0.0037	0.0068
			(0.0020)+	(0.0010)**	(0.0020)**
7		Log(Sales Rank)			0.0024
					(0.0002)**
8	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0282	-0.0302	
			(0.0026)**	(0.0011)**	
9		Popular Products (151-500 nationally)	0.0015	-0.001	
			(-0.0019)	(-0.0008)	
10		Moderately Popular Products (501-1500 nationally)	0.0005	-0.002	
			(-0.0016)	(0.0006)**	
11		Somewhat Less Popular Products (1501-5000 nationally)	-0.003	-0.0025	
			(-0.0026)	(0.0009)**	
12		Less Popular Products (5001-15000 nationally)	-0.0017	-0.0014	
			(-0.0019)	(0.0008)+	
13		Relative Price	0.0018	0.0029	0.0051
			(0.0057)	(0.0023)	(0.0059)
14		Log(Sales Rank)			0.0022
					(0.0004)**
15	Other	Relative price (not interacted)	-0.0226	-0.0237	-0.0136
			(0.0007)**	(0.0007)**	(0.0007)**
		Observations	4,051,254	4,051,254	4,051,254
		Number of FEs	978,611	978,611	978,611
		Within Estimator R-squared	0.07	0.07	0.07
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Book popularity spline (columns 1 & 2) -Log(Sales Rank) (column 3)	-Broadband competition -Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out)		

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects. + significant at 10%; * significant at 5%; ** significant at 1%. For sales rank regression, we use entry in a 5.4 miles radius.

Appendix Table 4: Alternate Entry (Only with a Two Month Lag) Results

Row			(1) 5.4 Miles	(2) 20 Miles	(3) Sales Rank
1	Discount Store Entry (interactions)	Very Popular Products[#] (Top 150 nationally)	-0.0292	-0.0348	
			(0.0012)**	(0.0007)**	
2		Popular Products (151-500 nationally)	0.002	-0.002	
			(0.0008)*	(0.0005)**	
3		Moderately Popular Products (501-1500 nationally)	-0.0034	-0.0064	
			(0.0006)**	(0.0003)**	
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.006	-0.0075	
			(0.0011)**	(0.0005)**	
5		Less Popular Products (5001-15000 nationally)	-0.0008	-0.0006	
			(0.0009)	(0.0004)	
6		Relative Price	0.0035	0.0044	0.0052
			(0.0022)	(0.0011)**	(0.0023)*
7		Log(Sales Rank)			0.0027
					(0.0002)**
8	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0288	-0.0309	
			(0.0033)**	(0.0014)**	
9		Popular Products (151-500 nationally)	0.0026	-0.0001	
			(0.0022)	(0.0011)	
10		Moderately Popular Products (501-1500 nationally)	0.0015	-0.0003	
			(0.0021)	(0.0009)	
11		Somewhat Less Popular Products (1501-5000 nationally)	-0.0027	-0.0012	
			(0.0033)	(0.0014)	
12		Less Popular Products (5001-15000 nationally)	-0.0023	-0.002	
			(0.0027)	(0.0011)+	
13		Relative Price	-0.0036	0.0071	0.0019
			(0.0071)	(0.0033)*	(0.0072)
14		Log(Sales Rank)			0.0023
					(0.0005)**
15	Other	Relative price	-0.0224	-0.023	-0.0134
			(0.0007)**	(0.0007)**	(0.0007)**
		Observations	4,051,254	4,051,254	4,051,254
		Number of FEs	978,611	978,611	978,611
		Within Estimator R-squared	0.07	0.07	0.07
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Book popularity spline (columns 1 & 2) -Log(Sales Rank) (column 3)	-Broadband competition -Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out)		

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects. + significant at 10%; * significant at 5%; ** significant at 1%. For sales rank regression, we use entry in a 5.4 miles radius.

Appendix Table 5: Alternate Entry (One Month Lead) Results

Row			(1) 5.4 Miles	(2) 20 Miles	(3) Sales Rank
1	Discount Store Entry (interactions)	Very Popular Products[#]	-0.0299	-0.0341	
		(Top 150 nationally)	(0.0011)**	(0.0007)**	
2		Popular Products	-0.0067	-0.0083	
		(151-500 nationally)	(0.0007)**	(0.0004)**	
3		Moderately Popular Products	-0.0079	-0.0088	
		(501-1500 nationally)	(0.0005)**	(0.0003)**	
4		Somewhat Less Popular Products	-0.0081	-0.0081	
		(1501-5000 nationally)	(0.0007)**	(0.0004)**	
5		Less Popular Products	-0.0017	-0.0018	
		(5001-15000 nationally)	(0.0005)**	(0.0002)**	
6		Relative Price	0.0084	0.0084	0.0077
			(0.0016)**	(0.0009)**	(0.0017)**
7		Log(Sales Rank)			0.0033
					(0.0002)**
8	Large Bookstore Entry (interactions)	Very Popular Products	-0.0338	-0.0321	
		(Top 150 nationally)	(0.0021)**	(0.0009)**	
9		Popular Products	-0.0076	-0.0075	
		(151-500 nationally)	(0.0015)**	(0.0007)**	
10		Moderately Popular Products	-0.0071	-0.0073	
		(501-1500 nationally)	(0.0011)**	(0.0004)**	
11		Somewhat Less Popular Products	-0.0098	-0.0071	
		(1501-5000 nationally)	(0.0017)**	(0.0006)**	
12		Less Popular Products	-0.0026	-0.0021	
		(5001-15000 nationally)	(0.0011)*	(0.0004)**	
13		Relative Price	0.0112	0.0076	0.0115
			(0.0041)**	(0.0014)**	(0.0043)**
14		Log(Sales Rank)			0.0038
					(0.0003)**
29	Other	Relative price	-0.0238	-0.0274	-0.0146
			(0.0008)**	(0.0008)**	(0.0007)**
		Observations	4,051,254	4,051,254	4,051,254
		Number of FEs	978,611	978,611	978,611
		Within estimator R-squared	0.07	0.07	0.07
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Book popularity spline (columns 1 & 2) -Log(Sales Rank) (column 3)	-Broadband competition -Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry with 1 month lag -Large Bookstore Entry with 1 month lag -Product-location fixed effects (differenced out)		

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects. + significant at 10%; * significant at 5%; ** significant at 1%. For sales rank regression, we use entry in a 5.4 miles radius.

Appendix Table 6: Alternative Broadband Definition: linear interpolation

Row			(1)	(2)	(3)
			5.4 Miles	20 Miles	Sales Rank
1	Discount Store Entry (interactions)	Very Popular Products[#]	-0.0320	-0.0372	
		(Top 150 nationally)	(0.0012)**	(0.0008)**	
2		Popular Products	-0.0034	-0.0061	
		(151-500 nationally)	(0.0008)**	(0.0005)**	
3		Moderately Popular Products	-0.0060	-0.0080	
		(501-1500 nationally)	(0.0006)**	(0.0003)**	
4		Somewhat Less Popular Products	-0.0082	-0.0084	
		(1501-5000 nationally)	(0.0009)**	(0.0004)**	
5		Less Popular Products	-0.0019	-0.0020	
		(5001-15000 nationally)	(0.0007)**	(0.0003)**	
6		Relative Price	0.0147	0.0107	0.0153
			(0.0022)**	(0.0010)**	(0.0022)**
7		Log(Sales Rank)			0.0034
					(0.0002)**
8	Large Bookstore Entry (interactions)	Very Popular Products	-0.0339	-0.0343	
		(Top 150 nationally)	(0.0025)**	(0.0011)**	
9		Popular Products	-0.0029	-0.0044	
		(151-500 nationally)	(0.0020)	(0.0008)**	
10		Moderately Popular Products	-0.0022	-0.0047	
		(501-1500 nationally)	(0.0016)	(0.0006)**	
11		Somewhat Less Popular Products	-0.0074	-0.0067	
		(1501-5000 nationally)	(0.0025)**	(0.0009)**	
12		Less Popular Products	-0.0023	-0.0022	
		(5001-15000 nationally)	(0.0018)	(0.0006)**	
13		Relative Price	0.0183	0.0145	0.0228
			(0.0061)**	(0.0021)**	(0.0063)**
14		Log(Sales Rank)			0.0035
					(0.0004)**
15	Other	Relative price	-0.0237	-0.0268	-0.0146
			(0.0007)**	(0.0008)**	(0.0007)**
		Observations	4,051,254	4,051,254	4,051,254
		Number of FEs	978,611	978,611	978,611
		Within estimator R-squared	0.06	0.06	0.06
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Broadband competition -Log(Number of reviews)	-Dummy for missing elapsed date information -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out) -Log(Sales Rank) (Column 3 only) -Book popularity spline (cols 1 & 2 only)		

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%. For sales rank regression, we use entry in a 5.4 miles radius.

Appendix Table 7: Differences in population growth based on census place data

Row			(1) High Growth	(2) Low Growth	(3) Intermediate Growth
1	Discount Store Entry (interactions)	Very Popular Products[#] (Top 150 nationally)	-0.0278	-0.0325	-0.0334
			(0.0028)**	(0.0022)**	(0.0018)**
2		Popular Products (151-500 nationally)	-0.0037	-0.0086	-0.0013
			(0.0020)+	(0.0015)**	(0.0012)
3		Moderately Popular Products (501-1500 nationally)	-0.0069	-0.0079	-0.005
			(0.0015)**	(0.0009)**	(0.0009)**
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0081	-0.0062	-0.0096
			(0.0025)**	(0.0012)**	(0.0013)**
5		Less Popular Products (5001-15000 nationally)	-0.0048	-0.0005	-0.0023
			(0.0016)**	(0.0008)	(0.0012)+
6		Relative Price	0.0106	0.0087	0.0208
			(0.0052)*	(0.0029)**	(0.0035)**
7	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0174	-0.0426	-0.0362
			(0.0053)**	(0.0051)**	(0.0034)**
8		Popular Products (151-500 nationally)	0.01	-0.004	-0.0056
			(0.0050)*	(0.0043)	(0.0024)*
9		Moderately Popular Products (501-1500 nationally)	0.0112	-0.0052	-0.0047
			(0.0044)*	(0.0033)	(0.0020)*
10		Somewhat Less Popular Products (1501-5000 nationally)	0.0001	-0.0043	-0.0097
			(0.0061)	(0.0055)	(0.0031)**
11		Less Popular Products (5001-15000 nationally)	0.0046	-0.006	-0.0025
			(0.0036)	(0.0043)	(0.0024)
12		Relative Price	0.0455	0.0467	0.0032
			(0.0192)*	(0.0155)**	(0.0067)
13	Other	Relative price	-0.0228	-0.0230	-0.0247
			(0.0016)**	(0.0016)**	(0.0011)**
		Observations	926,840	836,896	1,935,702
		Number of FEs	222,772	203,105	467,509
		Within estimator R-squared	0.07	0.07	0.08
		Controls -Dummy for missing price information -Average rating -Log(days since launch) -Time dummies -Broadband competition	-Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out) -Book popularity spline (cols 1 & 2 only)		

High growth locations are defined as those in which population change in the Census Place for the location between the 1990 and 2000 decennial Censuses was above the 75th percentile; low growth locations are defined as those for which the population change was below the 25th percentile; intermediate growth areas as defined as those between the 25th and 75th percentiles. 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 8: Difference ways to treat missing prices

Row			(1) Drop missing prices	(2) Includes missing price interaction	(3) Imputing prices using prior/future month prices	(4) Imputing prices using linear regression
1	Discount Store Entry (interactions)	Very Popular Products[#] (Top 150 nationally)	-0.0348	-0.0322	-0.0322	-0.0322
			(0.0013)**	(0.0012)**	(0.0012)**	(0.0012)**
2		Popular Products (151-500 nationally)	-0.0066	-0.0037	-0.0036	-0.0035
			(0.0009)**	(0.0008)**	(0.0008)**	(0.0008)**
3		Moderately Popular Products (501-1500 nationally)	-0.0086	-0.0062	-0.0063	-0.0063
			(0.0007)**	(0.0006)**	(0.0006)**	(0.0006)**
4		Somewhat Less Popular Products (1501-5000 nationally)	-0.0136	-0.0083	-0.0082	-0.0082
			(0.0011)**	(0.0009)**	(0.0009)**	(0.0009)**
5		Less Popular Products (5001-15000 nationally)	-0.0046	-0.0019	-0.0018	-0.0018
			(0.0007)**	(0.0007)**	(0.0007)**	(0.0007)**
6		Relative Price	0.0183	0.0157	0.0184	0.0175
			(0.0028)**	(0.0027)**	(0.0023)**	(0.0022)**
7		Missing Price		-0.0010		
				(0.0012)		
8	Large Bookstore Entry (interactions)	Very Popular Products (Top 150 nationally)	-0.0364	-0.0323	-0.0342	-0.0341
			(0.0028)**	(0.0026)**	(0.0025)**	(0.0025)**
9		Popular Products (151-500 nationally)	-0.0059	-0.0012	-0.0031	-0.0030
			(0.0022)**	(0.0021)	(0.0020)	(0.0020)
10		Moderately Popular Products (501-1500 nationally)	-0.0048	-0.0007	-0.0026	-0.0026
			(0.0019)*	(0.0018)	(0.0016)	(0.0016)**
11		Somewhat Less Popular Products (1501-5000 nationally)	-0.0146	-0.0064	-0.0076	-0.0075
			(0.0032)**	(0.0026)*	(0.0025)**	(0.0025)
12		Less Popular Products (5001-15000 nationally)	-0.0059	-0.0015	-0.0025	-0.0024
			(0.0023)**	(0.0019)**	(0.0018)	(0.0018)
13		Relative Price	0.0139	0.0111	0.0220	0.0210
			(0.0076)**	(0.0073)**	(0.0062)**	(0.0061)**
14		Missing Price		0.0073		
				(0.0028)**		
15	Other	Relative price (not interacted)	-0.0350	-0.0237	-0.0013	-0.0019
			(0.0009)**	(0.0008)**	(0.0006)*	(0.0005)**
		Observations	3,790,471	4,051,254	4,051,254	4,051,254
		Number of FEs	956,102	978,611	978,611	978,611
		Within Estimator R-squared	0.06	0.07	0.07	0.07
		Controls -Dummy for missing price (column 2) -Average rating -Log(days since launch) -Time dummies -Log(Sales Rank) (column 3) -Book popularity spline (columns 1, 2 & 4)	-Broadband competition -Dummy for missing elapsed date information -Log(Number of reviews) -Discount Store Entry -Large Bookstore Entry -Product-location fixed effects (differenced out)			

Robust standard errors are in parentheses and are clustered by location-time. Regressions include location-product fixed effects.
[#]Base is *unpopular books* ranked 15000 and up.
+ significant at 10%; * significant at 5%; ** significant at 1%. Entry is defined with a 5.4 mile radius.