Battle of the Retail Channels:
How Product Selection and Geography Drive Cross-channel Competition

June 2009

Erik Brynjolfsson*

MIT Sloan School of Management, erikb@mit.edu

Yu (Jeffrey) Hu

Krannert School of Management, Purdue University, yuhu@purdue.edu

Mohammad S. Rahman

Haskayne School of Business, University of Calgary, rahman@ucalgary.ca

Abstract

A key question for Internet commerce is the nature of competition with traditional brick-and-mortar retailers. Although traditional retailers vastly outsell Internet retailers in most product categories, research on Internet retailing has largely neglected this fundamental dimension of competition. Is cross-channel competition significant, and, if so, how and where can Internet retailers win this battle? This paper attempts to answer these questions using a unique combination of data sets. We collect data on local market structures for traditional retailers, and then match these data to a data set on consumer demand via two direct channels: Internet and catalog. Our analyses show that Internet retailers face significant competition from brick-and-mortar retailers when selling mainstream products, but are virtually immune from competition when selling niche products. Furthermore, since the Internet channel sells proportionately more niche products than the catalog channel, the competition between the Internet channel and local stores is less intense than the competition between the catalog channel and local stores. The methods we introduce can be used to analyze cross-channel competition in other product categories, and suggest that managers need to take into account the types of products they sell when assessing competitive strategies.

Keywords: Internet markets, electronic commerce, cross-channel, competition, retailing, channel, niche products, geography.

Acknowledgements: We thank Jason Abrevaya, Chris Anderson, Jack Barron, Prabuddha De, Alok Gupta, Jerry Hausman, Ramayya Krishnan, Barrie Nault, Vallabh Sambamurthy, three anonymous reviewers, and seminar participants at Purdue University, University of Calgary, the 2007 International Conference on Information Systems (ICIS), and the 2007 Big Ten Information Systems Symposium for valuable comments on this paper. We thank Zhulei Tang and Krannert Computing Center for providing computing resources. Generous funding was provided by the MIT Center for Digital Business and Purdue Research Foundation. Any remaining errors are solely the responsibility of the authors.

* The authors are listed in alphabetical order.

1. Introduction

> "Our primary competitors are brick-and-mortar, so we have to be really responsive from a fulfillment standpoint. More and more, we're going to be competing with the guy down the street where a customer can drive and pick up an order the same day."
>
> --*Kurt Goodwin, VP of Operations of Crutchfield, quoted in Dubbs (2002)*

A key question for Internet commerce is the nature of its competition with traditional brick-and-mortar commerce. In almost every product category sold on the Internet, consumers have the option of buying from traditional retailers instead. Although traditional retailers vastly outsell Internet retailers in almost every product category,[1] research on Internet commerce has largely neglected this fundamental dimension of competition. Do Internet retailers compete with traditional retailers? If they do, how and where can Internet retailers win this battle with traditional retailers?

Anecdotal evidence suggests that Internet retailers worry about the competition from traditional retailers. For instance, Jeff Wilke, SVP of Amazon, identified the "instant gratification" available in traditional stores as the key challenge to Amazon's growth.[2] To compete with traditional retailers, many Internet firms have built delivery centers all across the U.S., speeding up the delivery of their products to consumers. To attract consumers who usually do not pay shipping charges when purchasing from local stores, Internet retailers frequently offer free-shipping discounts. Furthermore, Internet retailers have invested heavily in technologies that allow consumers to inspect and sample products before purchasing, and in good customer services that alleviate consumers' concerns about returns and refunds.

Economic theory on firm competition has shown that as the number of firms in a market increases, the competition intensifies and each firm's market share drops (Tirole 1988). If one considers an Internet retailer as just another additional retailer, then such an Internet retailer seems likely to face stiff competition in a local market with many traditional retailers. However, what has been neglected by the literature on the competition between Internet and traditional retailers are Internet retailers' ability to

---

[1] ComScore, an online market research company, estimates that in 2006 the non-travel Internet retail has accounted for approximately 7 percent of U.S. consumers' retail spending excluding gas, autos and food (Rubin 2006).
[2] The authors' interview at Amazon Headquarters, Seattle, Washington, February, 2005.

dramatically reduce consumer transaction costs by making large product selections available and Internet retailers' ability to greatly reduce consumer search costs through the usage of IT-enabled tools. These costs have similar impacts on consumers' utilities; hence this paper uses a broad definition of "consumer search costs" to consider both types of costs. In contrast, consumers search costs in brick-and-mortar stores are low for only a few products that are widely available and highly visible.

We divide consumer demand into the demand for popular products and the demand for niche products. Because of the strong demand for popular products, it is an optimal strategy for brick-and-mortar stores to lower the search costs for these popular products, making them available and highly visible, and thus facilitating consumers' purchasing of them. In contrast, niche products are typically more difficult to find in traditional stores and may not be carried at all. Examples of popular products include bestselling books at bookstores, new-release movies at movie rental shops, Billboard top hits at music stores, and clothing stores' current-season products that follow the latest fashion trends. Compared with traditional stores, Internet retailers carry a wide selection of products and provide IT-enabled search tools, recommendation tools, and browsing functions, greatly lowering search costs for both popular and niche products. Therefore, an Internet retailer may face a lower level of competition from traditional stores when selling niche products that have high search costs in traditional stores, than when selling popular products.

The primary focus of our paper is to empirically study how the level of competition between Internet retailers and traditional stores varies across products that have different levels of search costs in traditional stores. We have collected a unique data set on local market structures – defined as the number of brick-and-mortar stores in a local market, and then matched it to a rich data set on consumer demand. We find that the impact of local market structures on consumers' Internet demand (as well as consumers' catalog demand) is negative and statistically highly significant for popular products. However, this impact is statistically insignificant for niche products.

Despite some similarities between the Internet channel and the catalog channel, previous research has shown that niche products can make up a larger percentage of a company's total sales through the Internet channel than through the catalog channel (Brynjolfsson et al. 2006). This paper brings to light an interesting implication of this phenomenon: since the Internet channel sells proportionately more niche products than the catalog channel, the competition between the Internet channel and local stores becomes less intense than the competition between the catalog channel and local stores.

Our paper has important managerial implications. First, the methods we introduce rely on readily accessible data on the number of local stores, and can be applied to analyze cross-channel competition in many other product categories. Second, our results suggest that Internet retailers are insulated from this competition, if they sell niche products that have high search costs in traditional stores. To mitigate the competition with traditional retailers, Internet retailers should adopt a strategy of making even more niche products available and developing more efficient IT-enabled search tools, recommendation systems, and etc. Furthermore, many Internet retailers use various measures to segment consumers and predict their future demand. Our findings suggest that local market structures – which can have an impact on consumers' Internet demand for popular products – should be used in the marketing decisions of Internet retailers, particularly those who derive a large share of their sales from popular products.

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature on this topic. In Section 3, we first discuss our data collection methodology and data sources. We then present our empirical analyses and results. In Section 4, the paper concludes with a discussion of our findings and some broader implications.

2. Literature Review

Theoretical research on the competition between direct retailing and traditional brick-and-mortar retailing can be traced to Balasubramanian (1998). By analyzing a game-theoretic model that has a direct retailer and multiple brick-and-mortar stores, he suggests that the direct retailer competes with local stores. In a landmark paper, Bresnahan and Reiss (1991) show that as the number of competing firms in a local market increases, the competition becomes more intensified and firms' profit margins fall. However, these papers do not study how the competition varies across different products with different search costs.

Previous empirical research has found that Internet markets can improve consumer welfare and firms' profits through wider product selection (Brynjolfsson et al. 2003, Cachon et al. 2007), and that consumer demand through the Internet channel is higher in local markets where local prices and sales tax rates are high (Goolsbee 2000, 2001, Chiou 2008, Ellison and Ellison 2006, Prince 2007). However, the impact of local market structures on consumers' Internet demand has not been explored by previous literature. In this paper, we aim to bridge this gap.

Our paper is closely related to two interesting papers that study how geographic variables have an impact on consumers' online behavior – Sinai and Waldfogel (2004), Forman et al. (2009), although it differs from them in many aspects. Sinai and Waldfogel (2004) find that consumers connect to the Internet to overcome spatial isolation (e.g., distance to retail stores, racial isolation). However, they do not study the competition between Internet retailing and brick-and-mortar retailing using actual demand data. Forman et al. (2009) find evidence that the existence of a discount store or a large bookstore in a geographic location decreases the likelihood that a popular book will appear in Amazon's list of top 10 bestselling books for that geographic location. Our paper differs significantly from their paper by studying how the number of local stores affects the individual-level demand rather than aggregate-level demand. Furthermore, we discuss how this effect varies for different channels and for different products.[3]

Recently, a few papers in marketing have leveraged spatial data to understand consumers' behaviors (see Bradlow et al. 2005 for a review). For example, Jank and Kannan (2005) show that including spatial dependence can help predict whether a consumer purchases an electronic copy or a print copy of the same book. We contribute to this nascent literature by highlighting how the local market structure, which varies with geographic location, can affect a consumer's purchasing behaviors on the Internet.

Next, we present our empirical analyses. To provide a richer understanding of the mitigating role played by search costs, we begin by studying the impact of local market structures on consumers' Internet demand. We then investigate how the effect varies across different products and across different channels.

3. Data and Empirical Analyses

[3] We directly measure individual-level demand, while they use sales rank for top products at geographic locations to make inferences about demand. We directly measure the number of physical stores at the zip code level, while they use the existence of stores at aggregate geographic locations which include large cities and small towns.

For this study, we have collected data from several sources. Our consumer demand data comes from a large retailer of women's clothing products.[4] The retailer primarily operates in two channels: the Internet channel (website) and the catalog channel (mail and telephone orders), with both channels contributing almost equally to the firm's revenue. We have information regarding all transactions made from May 19, 2003 to June 15, 2006 through these two primary channels.[5] For each item purchased from the retailer, we have information regarding the price paid, date of transaction, consumer's unique identification number, channel used to purchase the item, and transaction identification number. Overall we have records of about 7 million transactions that were made by approximately 1 million unique consumers. Moreover, we have information regarding what catalogs and emails each consumer received between January 2005 and June 2006. We also know each consumer's home zip code. This unique dataset enables us to determine both the overall demand and channel-specific demand at the direct retailer for each consumer. The data also allows us to determine the local market structure for each consumer.

An important feature of the retailer is that it offers exactly the same product selection (and prices) through its Internet and catalog channels. This eases the firm's logistic and ordering processes. In addition, the firm uses the same order fulfillment methods and facilities for the two channels. These decisions greatly facilitate our research design by automatically controlling for differences in sales tax policies, shipping costs, and the possibility of stock outs, eliminating these alternative explanations for potential differences in the demand across the two channels.

Our data on local market structures comes from Superpages.com (http://www.superpages.com), a Verizon spin-off that is a prominent provider of yellow pages and information services. We have obtained a comprehensive list of 41,513 zip codes served by the US Postal Service as of November 2006 in the 50 U.S. states and the District of Columbia. We have then employed a set of web-crawling programs to collect data from this website. For each zip code, we have collected the number of women's clothing

[4] The retailer requests to remain anonymous.

[5] The retailer also has a physical store that accounts for a negligible percentage of overall sales. We do not have any information regarding the transactions made in the physical store. Our results are robust to including or excluding consumers who have access to the physical store.

stores listed on Superpages.com that are within 5 miles (as well as 10, 15, 20, 25, and 50 miles) from the center of that zip code. To ensure data consistency, we have collected this data three times in October and November of 2006. The differences among the data sets collected in these three snapshots are negligible. We use the Superpages data from the last collection. To further test data accuracy, we have collected data from Yahoo Local (http://local.yahoo.com), a leading online portal that provides the number of women's clothing stores within a certain radius from the center of each zip code. We have found an extremely high Pearson correlation coefficient between Superpages data and Yahoo data. We use Superpages data in this paper, and we note that as expected, all the results remain unchanged if Yahoo data is used.

In our analyses, we study consumer demand in 2006 (January 1, 2006—June 15, 2006), the latest year for which we have data. An important consideration in selecting the sample for our study is to control for the effect of advertising on consumers' demand. The retailer mainly promotes its products by sending catalogs. Our analysis of the data shows that the impact of a catalog typically lasts for about 30 days. This is consistent with the retailer's past experience which suggests that the effect of a catalog lasts for 30-45 days. To be more conservative, we only include U.S. consumers who have received all the catalogs that were sent out between November 1, 2005 (61 days prior to January 1, 2006) and June 15, 2006. In addition, since the retailer has a physical store in Florida, it needs to collect sales tax on sales to Florida consumers. Although our results remain practically unchanged even if we include Florida consumers in our analyses, we have excluded all the Florida consumers to eliminate the difference in sales taxes as a confounding factor. Correspondingly, we have retained 163,933 consumers for our analyses.

The sample of consumers we analyze is quite representative of the U.S. population, as evidenced by the fact that these consumers live in all 50 states and 20,005 zip codes (out of 41,513 zip codes). The mean of the number of local stores within 5 miles is 30 for our sample; for the U.S. population, the mean is 29. Table 1 presents the descriptive statistics of individual-level demand. During the period between January 1, 2006 and June 15, 2006, 21% of the consumers in our sample had positive demand and the average number of items bought from the retailer is 0.94 for the consumers in our sample, with approximately 47% of the consumers' purchases occurring through the Internet channel.

Table 1: Descriptive Statistics of Consumers' Demand

	Mean	Std. Deviation	Min	Max
Internet Channel	0.44	1.90	0	87
Catalog Channel	0.50	1.87	0	78
Overall	0.94	2.68	0	109

For each consumer's home zip code, we collect demographic and socioeconomic variables from U.S. Census 2000. These variables are used as control variables in our analyses. Since our analyses focus on transactions occurred in 2006, the earlier transaction data (May 2003 – December 2005) is used to calculate historical purchasing measures, which are then used as controls for consumer heterogeneity.

3.1. Effect of Local Market Structure on Internet Channel

3.1.1 Econometric Models

We first use a probit model to study whether the local market structure has an impact on the propensity of purchasing from the Internet retailer. Assuming I_i is an indicator of whether consumer *i* had positive demand in 2006, and X_i is a vector of explanatory variables, we can estimate the following probit model:

$$P(I_i = 1 \mid X_i) = \Phi(X_i \beta) . \tag{1}$$

Next, we investigate the impact of the local market structure on the variation in overall demand on the Internet. Note that the number of items purchased, which is count data, can be assumed to follow a Poisson distribution. Since we find evidence of over-dispersion, we estimate a negative binomial regression model, which is a generalization of a Poisson regression model that allows for over-dispersion by incorporating an individual unobserved effect into the conditional mean (Hausman et al. 1984):

$$f(y_i \mid X_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i !}, y_i = 0,1,2,3,... \tag{2}$$

where: y_i is the number of items purchased by consumer *i*; X_i is a vector of explanatory variables; $E(y_i \mid X_i) = \mu_i = \exp(X_i \beta + \varepsilon_i)$ is the conditional mean; ε_i is the unobserved heterogeneity and is assumed to follow a log-gamma distribution, with $\varepsilon_i \sim \Gamma(\theta, \theta)$ (Cameron and Trivedi 1998). One can test for over-dispersion by testing the hypothesis $\theta = 0$.

3.1.2. Control Variables

A consumer's demographic and socioeconomic variables such as her income, age, education, and gender may influence her demand (Goolsbee 2000, 2001). Therefore, we use demographic and socioeconomic variables collected from U.S. Census 2000 as control variables. Accordingly, control variables in our analyses include the natural log of median household income, percentage of females in the population, percentage of the population with at least a Bachelor's degree, and median age of the female population, all at the level of the consumer's home zip code.[6] In addition, whether a consumer lives in an urban area or not may influence that consumers' demand (Glaeser et al. 2001). We include a population density variable, defined as the population per square mile divided by 10,000, as a control variable.[7]

Historical purchasing measures ("Recency", "Frequency", and "Monetary Value") are widely used in the retailing industry as well as in the academic literature to segment consumers into loyal and non-loyal consumers and to control for consumer heterogeneity (e.g., Anderson and Simester 2004). "Recency" is commonly defined as the number of periods since the last purchase; "Frequency" is defined as the total number of items ordered over a time period; and "Monetary Value" is defined as the average per-item price a consumer paid over a time period. We use the natural log of these three measures, calculated using data from May 2003 to December 2005, as control variables for consumer heterogeneity.[8]

3.1.3. Results

We let the variable *NumStores* be the natural log of the number of local stores within 5 miles of each consumer's home zip code.[9] In columns (1) and (2) of Table 2, we present the estimates obtained from the probit model and the negative binomial model respectively. Our coefficient of interest, i.e., the coefficient of *NumStores*, in both models is negative but barely significant at the 5% significance level. This indicates that an increase in local stores does decrease the demand from the Internet, but the overall effect

---

[6] Since U.S. Census demographic and socioeconomic variables are not available for several zip codes, controlling for demographic and socioeconomic variables slightly reduces our sample size to 163,891.

[7] We have calculated the correlation coefficients among control variables and the variable *NumStores* and find none of these correlation coefficients is high enough to cause multicollinearity concerns.

[8] Descriptive statistics of all the independent variables are provided in Table A1 of Appendix A.

[9] We add one to the number of local stores before taking the natural log. Our results are robust to using different constants (e.g., c = 0.1, 0.2, 0.5, 2) in the logarithmic transformation log(c+x).

is relatively small. For instance, *ceteris paribus*, a consumer with 7 (median) local stores nearby has an overall Internet demand that is 2.5% less than a consumer with zero (25th percentile) stores nearby.

Table 2: Effect of Local Market Structure on Internet Demand

	Probit (1)	Negative Binomial (2)	IV Probit[a] (3)	IV Negative Binomial[b] (4)
NumStores	-0.007*	-0.012*	-0.007	-0.013
	(0.003)	(0.006)	(0.004)	(0.008)
Recency	-0.152**	-0.254**	-0.152**	-0.254**
	(0.005)	(0.007)	(0.005)	(0.007)
Frequency	0.189**	0.375**	0.189**	0.375**
	(0.006)	(0.011)	(0.006)	(0.011)
Monetary Value	-0.063**	-0.106**	-0.063**	-0.107**
	(0.006)	(0.011)	(0.006)	(0.012)
Median Income	0.135**	0.253**	0.131**	0.245**
	(0.019)	(0.034)	(0.019)	(0.035)
Population Density	-0.020*	-0.039*	-0.020*	-0.038*
	(0.008)	(0.016)	(0.009)	(0.019)
Median Age of Female	-0.003**	-0.006**	-0.003**	-0.005**
	(0.001)	(0.002)	(0.001)	(0.002)
Percentage with Bachelor's Degree	0.160**	0.296**	0.161**	0.302**
	(0.045)	(0.081)	(0.047)	(0.085)
Percentage Female Population	-0.338	-0.708	-0.332	-0.689
	(0.230)	(0.414)	(0.234)	(0.429)
Intercept	-1.792**	-2.010**	-1.755**	-1.945**
	(0.243)	(0.440)	(0.245)	(0.449)
Log Likelihood	-51666.37	-91863.45		-90517.65
Sample Size	163,891	163,891	161,856	161,856

Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$. * $p < 0.05$.

[a] The chi-square model fit statistics is 3387.10, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

3.2. Addressing Endogeneity

There can be concerns regarding unobservable factors that affect both the number of local stores and consumer's Internet demand. For example, suppose that local stores exit the market due to local predisposition toward Internet retailers. This would reverse the direction of causality in our estimates. More subtly, suppose that there is some unobserved socioeconomic factor or preference that leads to a higher Internet demand while affecting the number of local stores. Therefore, we need to address the endogeneity concern ensued from potential simultaneity and missing variable problems.

To address this concern, we collect data on the number of local stores in 1994 (*NumStores94*) and use it as the instrument variable for the number of local stores (*NumStores*).[10] The number of local stores in 1994 is correlated with the number of local stores in 2006. However, since the number of local stores in 1994 predates the rise of the ecommerce, it should be uncorrelated with the unobservable factor or preference that affects consumer's Internet demand in 2006.

We address the endogeneity concern in both the probit and negative binomial models. In the probit model we use Newey (1987)'s minimum chi-squared estimator. The estimates from the instrumented probit model are presented in column (3) of Table 2. The chi-square value associated with the Wald test of endogeneity is very small (0.02), and fails to reject the null hypothesis that *NumStores* is exogenous. In the negative binomial model, we follow the two-stage approach described in Mullahy (1997). In the first stage, we calculate the predicted *NumStores* by regressing *NumStores* onto all the control variables and *NumStores94*. Next, we include the predicted value of *NumStores* when estimating the negative binomial model in the second stage. The estimates are presented in column (4) of Table 2. The Hausman specification test for endogeneity using the coefficients in column (4) of Table 2 produces a t statistic of -0.194, indicating that we fail to reject the null hypothesis that *NumStores* is not endogenous. Reassuringly, non-IV and IV estimators produce similar coefficient estimates, as shown in Table 2. This provides further evidence that our results are robust to using instrumental variable estimators. More details on how we test for endogeneity and address endogeneity concerns can be found in Appendix C.

3.3. Effect of Local Market Structure Varies across Different Products

In the clothing industry, companies closely monitor market trends by attending runway shows, reading trade journals, obtaining market-trend reports, and shopping in the market (Rantisi 2002). When a

[10] We collect the number of local women's clothing stores in 1994 from the U.S. Census ZIP Code Business Patterns (http://www.census.gov/mp/www/cat/business_and_industry/006338.html). We use ArcGIS 9 to calculate the number of stores within 5 miles of each zip code. Using this instrumental variable reduces our sample size to 161,856.

particular product or design catches on, it would be quickly copied by many companies — each company would offer its own version of the products or designs that are popular (Varian 2007, Raustiala and Sprigman 2006). The phenomenon of brick-and-mortar retailers focusing on popular products is not surprising: popular products account for the bulk of their sales and they may not carry a large selection of niche products. It is a well-known result in the operations management literature's "newsvendor model" that a product's inventory level and whether it is carried depend on the product's demand rate (Cachon and Terwiesch 2008). Thus, it is not profitable for a retailer to carry even one unit of a product if its demand rate is too low. The marketing literature also shows that retailers typically cut the low-selling SKUs first when trying to create an "efficient assortment" and lower their costs (Broniarczyk et al. 1998).

In contrast, Internet retailers incur dramatically lowered inventory costs because they sell from a few centralized warehouses to consumers living across the whole country. As a result, they can offer a broad product selection that includes both popular products and niche products. According to our discussion with the senior management of the focal retailer, this retailer offers popular products that follow market trends and have large sales; this retailer also offers many niche products that have small sales.

It is reasonable to expect that the difficulty of finding a substitute for a certain product in brick-and-mortar stores is correlated with the product's popularity. That is, it is easier for consumers to find a substitute in local stores for an Internet retailer's popular products than for its niche products. This is because products that can substitute for niche products are either unavailable (due to the limit of shelf space and high inventory cost) or difficult to find even when they are available (due to the difficulty of searching in brick-and-mortar stores). Therefore, we posit that the competition between an Internet retailer and local stores would be intense for popular products and that an Internet retailer is likely to face little or no competition when selling niche products. Next, we empirically test these predictions.

In our sample, 1,866 unique products have positive sales. We rank the sales of all 1,866 products to identify the top bestselling products. In the spirit of the classical "80/20 Rule", we identify the top bestselling products that cumulatively generate 80% of total sales as "popular products", leading to 253 popular products with average per-product sales of 534. The rest of the products are identified as "niche products", with average per-product sales of 21.[11] In our sample, 31,630 consumers have positive demand for popular products, and 14,149 consumers have positive demand for niche products. We note that, because of the well-studied market trends in the clothing industry, there exists a strong correlation between the focal retailer's product being "niche" and the difficulty of finding a substitute for it in brick-and-mortar stores. Our hypotheses are built on this correlation, rather than the absolute unavailability of the retailer's "niche products" in any brick-and-mortar stores. Our interpretation is valid as long as, on average, it is more difficult to find substitutes in local stores for an Internet retailer's "niche product" than for its "popular products". We also note that our results are robust to using different ways of identifying popular products: for instance, our results remain qualitative unchanged when we identify popular products as the top bestselling products that cumulatively generate 50%, 70%, or 90% of total sales.

We estimate the negative binomial model in Equation (2), with the Internet demand for popular products and the Internet demand for niche products as the dependent variable, respectively. We present these results in Columns (2) and (3) of Table 3.[12] For popular products, the coefficient on *NumStores* is negative and statistically significant, indicating that the Internet demand for popular products declines as the number of local stores increases. To further interpret this coefficient, we calculate how consumers' Internet demand changes when the number of local stores changes. Holding all the control variables in our analysis constant, a consumer with 7 (the median) local stores nearby has an Internet demand for popular products that is 4.2% less than that of a consumer with zero (the 25th percentile) store nearby.[13] In contrast, for niche products, the coefficient on *NumStores* is statistically insignificant. Thus, the impact of local stores on consumers' Internet demand is almost entirely via popular products. Meanwhile, niche products offered by the Internet channel are virtually immune from the cross-channel competition.

---

[11] We recruited an independent shopper and gave her a list of ten products (five from each category) and their pictures. We did not tell her which are "popular products" and which are "niche products", and asked her to shop for substitutes for these ten products in the local area of West Lafayette, IN. The shopper visited 10 local clothing stores and reported the number of local stores selling a similar product for each of the ten products. The number of local stores selling a substitute is 10, 4, 0, 4, and 6 for popular products, and is zero for all five niche products.

[12] Appendix B presents estimates from the probit model for this table and other tables, along with additional information.

[13] Table A2 in Appendix A presents reductions in Internet demand for popular products of a consumer with different number of stores nearby compared to a consumer with zero stores.

Table 3: Effect of Local Market Structure on Consumer Demand

	Internet Channel			Catalog Channel		
	Overall (1)	Popular (2)	Niche (3)	Overall (4)	Popular (5)	Niche (6)
NumStores	-0.012*	-0.020**	0.006	-0.024**	-0.032**	0.010
	(0.006)	(0.006)	(0.009)	(0.005)	(0.006)	(0.009)
Recency	-0.254**	-0.234**	-0.327**	-0.242**	-0.213**	-0.341**
	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.011)
Frequency	0.375**	0.387**	0.566**	0.225**	0.232**	0.399**
	(0.011)	(0.011)	(0.015)	(0.009)	(0.010)	(0.016)
Monetary Value	-0.106**	-0.102**	-0.247**	-0.070**	-0.069**	-0.179**
	(0.011)	(0.012)	(0.017)	(0.009)	(0.010)	(0.017)
Median Income	0.253**	0.268**	0.237**	0.074*	0.104**	-0.071
	(0.034)	(0.036)	(0.051)	(0.031)	(0.032)	(0.051)
Population Density	-0.039*	-0.036*	0.013	0.022	0.016	0.049**
	(0.016)	(0.017)	(0.019)	(0.012)	(0.013)	(0.018)
Median Age of Female	-0.006**	-0.005*	-0.008**	0.013**	0.014**	0.005
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Percentage with Bachelor's Degree	0.296**	0.346**	0.120	-0.253**	-0.251**	-0.292*
	(0.081)	(0.085)	(0.120)	(0.075)	(0.079)	(0.126)
Percentage Female Population	-0.708	-0.820	-1.329*	0.416	0.507	0.343
	(0.414)	(0.434)	(0.588)	(0.374)	(0.394)	(0.626)
Intercept	-2.010**	-2.577**	-2.464**	-0.937*	-1.692**	-0.461
	(0.440)	(0.463)	(0.640)	(0.397)	(0.419)	(0.659)
Log Likelihood	-91863	-81328	-40581	-109077	-99787	-37217
Sample Size	163,891	163,891	163,891	163,891	163,891	163,891

Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$; * $p < 0.05$.

3.4. Effect of Local Market Structure Varies across Different Channels

In many respects, the Internet channel is similar to the catalog channel. However, previous literature has found that niche products can make up a larger percentage of a company's total sales through the Internet channel than through the catalog channel (Brynjolfsson et al. 2006). Our results from Section 3.3 show that the competition between Internet retailing and brick-and-mortar retailing varies for different products; consequently, it is likely that the proportionally higher demand for niche products on the Internet mitigates the Internet channel's competition with local stores.

To provide richer insights regarding the mitigating role played by search costs and the relatively higher demand for niche products, we conduct our analyses in three steps. First, we analyze whether the Internet channel sells proportionately more niche products than the catalog channel. Second, we study whether the effect of the local market structure varies for different products on the catalog channel. Finally, we use an equation to show that the effect of the local market structure on overall demand can be written as the weighted average of such an effect on the demand for popular products and such an effect on the demand for niche products. This equation illustrates why the relatively higher demand for niche products on the Internet can mitigate the Internet channel's competition with brick-and-mortar stores.

First, we compare the concentration of product sales through the Internet channel with that through the catalog channel. Using the same definition of "popular products" and "niche products" as in Section 3.3, we find that niche products account for 24.4% of total sales through the Internet channel and 15.8% of total sales through the catalog channel. We interpret this as evidence that the Internet channel sells proportionately more niche products than the catalog channel.[14]

Next we analyze the competition between the catalog channel and local stores. Once again, we use a negative binomial model to estimate the impact on the overall catalog demand and the catalog demand for different products. These results are reported in columns (4)-(6) of Table 3. Interestingly, the results show that the coefficient of *NumStores* is negative and highly significant, when the overall catalog demand is used as the dependent variable. This suggests that the catalog channel strongly competes with brick-and-mortar stores, while the Internet channel weakly competes with local stores. In particular, the coefficient for the Internet channel, which is -0.012, is smaller in size than the coefficient for the catalog channel, which is -0.024. We interpret this as evidence that the competition between the Internet channel and local stores is less intense than the competition between the catalog channel and local stores.

When the catalog demand for popular products is used as the dependent variable, the coefficient of *NumStores* is negative and highly significant. This indicates that the catalog channel, like the Internet channel, competes with brick-and-mortar stores when selling popular products. In contrast, when the

[14] If we estimate a log-linear relationship (the Pareto curve regression used by Brynjolfsson et al. 2003) between sales and sales rank, we find that the slope parameter in that Pareto curve regression is -1.57 for the Internet channel and -1.81 for the catalog channel. A t-test shows that the difference between these two slope parameters is highly significant, indicating a less concentrated sales distribution for the Internet channel than for the catalog channel.

catalog demand for niche products is used as the dependent variable, the coefficient of *NumStores* is statistically insignificant. This suggests that the catalog channel, like the Internet channel, does not compete with brick-and-mortar stores when selling niche products.

Finally, we use an equation to formally illustrate the intuition that the Internet channel, compared with the catalog channel, can be relatively less affected by the local market structure because of the proportionately higher demand for niche products on the Internet. The coefficient of *NumStores* in Equation (2) can be expressed as the marginal effect of *NumStores* on the natural log of total demand:

$$\beta = \frac{\partial \ln E(y_i \mid X_i)}{\partial x_i} = \frac{\partial E(y_i \mid X_i)}{\partial x_i} \frac{1}{E(y_i \mid X_i)}, \tag{3}$$

where y_i is the total demand, X_i is a vector of explanatory variables, x_i is the local market structure variable, and β is the effect of the local market structure variable on the total demand.

The total demand can be written as the sum of the demand for popular products (y_{iP}) and the demand for niche products (y_{iN}). Substituting $y_i = y_{iP} + y_{iN}$ into (3) gives us

$$\beta = \frac{\partial(E(y_{iP} \mid X_i) + E(y_{iN} \mid X_i))}{\partial x_i} \frac{1}{E(y_i \mid X_i)} = \beta_P \frac{E(y_{iP} \mid X_i)}{E(y_i \mid X_i)} + \beta_N \frac{E(y_{iN} \mid X_i)}{E(y_i \mid X_i)}, \tag{4}$$

where β_P is the effect on the demand for popular products and β_N is the effect on the demand for niche products.

Results in Table 3 show that, for both the Internet and the catalog channels, the effect of local stores on the demand for popular products is negative and significant, while the effect of local stores on the demand for niche products is insignificant. In addition, niche products account for a larger proportion of total demand through the Internet channel (24.4%) than through the catalog channel (15.8%). Therefore, using Equation (4), it is easy to show that the effect of local stores on the total demand should be smaller in size for the Internet channel than for the catalog channel. By greatly lowering search costs, the Internet is not only flattening the sales distribution, but also mitigating the competition with local stores.

3.5. Robustness Check

There might be concerns regarding the competition from large local stores such as Wal-Mart and Target, which are not women's clothing stores but may sell some women's clothing products. We need to check the robustness of our results by considering the presence of Wal-Mart and/or Target stores in some local areas. We have collected the number of Wal-Mart and Target stores available within 5 miles of every zip code.[15] Our data shows that 58% of consumers in our sample have access to at least one Wal-Mart store and 41% of consumers in the sample have access to at least one Target store. Next, we have recoded *NumStores* by adding the number of Wal-Mart and Target stores to the original *NumStores*. We then re-estimate the impact of recoded *NumStores* on the overall demand as well as the demand for popular and niche products through the Internet and catalog channels, while keeping all control variables intact. The results of estimating the (non-IV) negative binomial regression model in Equation (2) are reported in Table D1 and D2 of Appendix D. Reassuringly, the estimates are qualitatively similar to the results presented earlier. Thus, our results are robust to including Wal-Mart and Target stores in *Numstores*.

There are varieties of other ways to further test the robustness of our results. One approach is to estimate the effect of local market structures on consumer demand using dummy variables instead of a continuous measure. We create a dummy variable *StoreAbove0* indicating whether there is at least one store within 5 miles, and a dummy variable *StoreAboveMedian* indicating whether there are more than 7 stores (which is the median of the distribution of the number of stores) within 5 miles. We have confirmed the robustness of the results presented in Table 3 by using these dummy variables. In addition, we have repeated our analyses by estimating the probit (or logit) model with these dummy variables (instead of estimating the negative binomial model), by using continuous measure of the number of stores within 10 miles, 15 miles, 25 miles, and 50 miles (instead of within 5 miles).[16] Consistently we have found a qualitatively similar result: the local market structure has a negative impact on Internet demand for popular products and an insignificant impact on Internet demand for niche products. Finally, all of our

[15] Target.com does not directly show the number of stores within a certain radius. However, it lists the addresses of all Target stores. We use ArcGIS 9 to calculate the number of stores within 5 miles of each zip code.

[16] One could define an area as "being an urban area" if the area's population density is above a certain threshold, or if the area's zip code lies within one of the 18 Consolidated Metropolitan Statistical Areas (CMSA). Our results are also robust to using such a dummy variable as a control variable or clustering standard errors by CMSA.

results remain essentially unchanged if we use the data from Yahoo Local, instead of the data from Superpages.com. All of these additional robustness results are available upon request.

3.6. Discussion

In this study, we integrate the theories of search costs and competition to investigate the nature of the competition between Internet and traditional retailers. Our empirical analyses emphasize the mitigating role of search costs in this competition. Our findings advance the existing literature that has primarily emphasized how electronic markets can improve consumer welfare through lower prices (Brynjolfsson and Smith 2000, Clay et al. 2002, Clemons et al. 2002), improved product variety (Brynjolfsson et al. 2003), and lower search costs (Bakos 1997). We find strong evidence that the greatly lowered search costs on the Internet not only flattens the sales distribution but also mitigates the competition between Internet and traditional retailers. More specifically, we compare two direct retailing channels – the catalog and the Internet – with identical product offerings and order fulfillment methods. Still, we find that an increase in local stores significantly reduces the catalog demand, whereas the impact of local stores on the Internet demand is smaller, a difference that can be attributed to lower search costs on the Internet.

In addition, our results are economically significant and meaningful. For example, everything else being equal, a consumer with 7 (median) physical clothing stores nearby has an Internet demand for popular products that is 4.2% less than that of a consumer with zero (25th percentile) store nearby. In comparison, there would need to be a 15.5% reduction in income to achieve the same level of change (4.2%) in consumers' Internet demand. Thus, the amount of local competition is an economically significant determinant of consumer's Internet demand for popular products. On the other hand, the impact of local market structures on consumers' Internet demand for niche products is negligible. This interesting dichotomy underscores the economic importance of product selection and search costs in shaping the competition between Internet and traditional commerce.

Our findings may have broad applicability. Our data comes from a women's clothing retailer that serves the general population of U.S. consumers in all 50 states. Consumers in our sample are quite similar to the U.S. population as measured by the number of local stores nearby. For example, 24% of the U.S.

population has no women's clothing stores within 5 miles as compared to 27% of consumers in our sample; 45% of the U.S. population has access to less than 7 stores (the median of the number of stores in our sample). Similarly, our sample compares well with the U.S. population in socioeconomic characteristics. For instance, the median household income of the U.S. population is $41,000, while the median household income of our sample is $51,700.

We have studied one product category of Internet retailing, but we expect similar results could be found for other categories, although the magnitude of the effect may vary. In fact, the effects may be stronger in certain other product categories such as books, music, and DVDs where products have unique identifiers. Unique identifiers in these categories should make it easier to identify an exact offline substitute for an online product, intensifying the competition between Internet and traditional retailers. Thus, we expect to see a larger impact of local market structures on Internet demand for popular products in categories such as books, music, and DVDs than in the clothing category. On the other hand, we expect offline search costs to remain high for niche products in these categories. As a result, the competition between Internet and traditional retailers would be muted for niche products. Perhaps more importantly, our model can be readily applied to any product category with varying levels of consumer search costs across products. The methods we introduce in this paper rely on readily accessible data – current and previous data on the number of physical stores near any given customer address, and the econometric techniques are quite general. Other researchers and managers should be able to quantify the extent of competition, or lack thereof, between retail channels for any type of products for which they have sales data.

4. Conclusions

While it is widely believed that Internet commerce competes strongly with traditional brick-and-mortar commerce, we demonstrate that the competition between Internet retailers and brick-and-mortar stores can be mitigated by the high consumer search costs in brick-and-mortar stores. Information technology in general and Internet markets in particular greatly lowers consumes search costs, allowing consumers to search beyond a few popular products and discover niche products. In contrast, consumers in brick-and-mortar stores often focus on a few popular products that are highly visible and do not discover a lot of niche products that have high search costs, even if these niche products are available. Therefore, when consumers purchase niche products, they do not view brick-and-mortar stores as good substitutes for Internet retailers. But the competition between these two types of retailers can be intense when consumers purchase popular products that have low search costs in both Internet and traditional stores.

Our results have important and direct implications for managers of Internet retailers. To date, a typical response of Internet retailers to the competition from brick-and-mortar stores has always been "catching up to local stores". Thus, Internet firms have invested heavily in delivery centers, customer services, and free shipping discounts. However, our paper points to a different strategy for Internet retailers to win this battle with local stores – differentiating from local stores by facilitating the purchasing of niche products that have high search costs in traditional stores. To mitigate the competition with local stores and accelerate their own revenue growth, Internet retailers can stock even more niche products, and, at the same time, help consumers discover and purchase them, either by deploying more effective IT-enabled tools or by training customer service representatives in helping consumers locate niche products more efficiently. Not surprisingly, some Internet retailers have started to explore this "niche" strategy to fuel their growth. Tony Hsieh, CEO of the largest Internet shoe retailer Zappos, said, "many consumers came to the company's website looking for offbeat styles. The more variety the company put online, the faster it grew. Today the company sells more than three million products across 1,000 brands" (Greco 2007).

Furthermore, variables that can help predict consumers' future demand are vital to the marketing plans of firms in the retailing industry. To date, Internet (and catalog) retailers have been using various measures of consumers' historical purchases to predict future demand. This paper shows that local market structures can have an impact on consumers' Internet demand (and catalog demand), and this impact is particularly strong for popular products. Therefore, local market structure variables could be included in the marketing decisions of Internet (and catalog) retailers, particularly those who derive a large proportion of their revenues from popular products. For instance, Internet retailers could vary their promotional strategies and product offerings based on the geographic location of the consumer.

Information technologies in general and Internet markets in particular have lowered consumer search costs. As a result, niche products account for a large percentage of Internet sales. We find that niche products in the Internet channel face little competition from local stores. Managers in either channel who recognize this fact may seek to increase this dimension of differentiation to further limit competition, and thereby increase profits in both channels.

References

Anderson, E., D. Simester. 2004. Long-run effects of promotion depth on new versus established customers: Three field studies. *Marketing Science* **23**(1) 4-20.

Bakos, J.Y. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science* **43**(12) 1676-1692.

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

Bradlow, E., B. Bronnenberg, G. Russell, N. Arora, D. Bell, S. Duvvuri, F. Hofstede, C. Sismeiro, R. Thomadsen, S. Yang. 2005. Spatial models in marketing. *Marketing Letters* **16**(3/4) 267-278.

Bresnahan, T.F., P.C. Reiss. 1991. Entry and competition in concentrated markets. *Journal of Political Economy* **99**(5) 977-1009.

Broniarczyk, S., W. Hoyer, L. McAlister. 1998. Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. Journal of Marketing Research 35(2) 166-176.

Brynjolfsson, E., Y.J. Hu, D. Simester. 2006. Goodbye pareto principle, hello long tail: The effect of internet commerce on the concentration of product sales. Purdue University.

Brynjolfsson, E., Y.J. Hu, M.D. Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* **49**(11) 1580-1596.

Brynjolfsson, E., M.D. Smith. 2000. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science* **46**(4) 563-585.

Cachon, G., C. Terwiesch. 2008. Matching supply with demand: An introduction to operations management. McGraw-Hill, Boston, MA.

Cachon, G.P., C. Terwiesch, Y. Xu. 2007. On the effects of consumer search and firm entry in a multiproduct competitive market. *Marketing Science* forthcoming.

Cameron, A.C., P.K. Trivedi. 1998. *Regression analysis of count data*. Cambridge University Press, Cambridge.

Chiou, L. 2008. Empirical analysis of competition between Wal-Mart and other retail channels", *Journal of Economics and Management Strategy*, forthcoming.

Clay, K., R. Krishnan, E. Wolff, D. Fernandes. 2002. Retail strategies on the web: Price and non-price competition in the online book industry. *Journal of Industrial Economics* **50**(3) 351-367.

Clemons, E.K., I.H. Hann, L.M. Hitt. 2002. Price dispersion and differentiation in online travel: An empirical investigation. *Management Science* **48**(4) 534-549.

Dubbs, D. 2002. Mind games. *Multichannel Merchant Magazine*, Stamford, CT (January 1).

Ellison, G.D., S.F. Ellison. 2006. Internet retail demand: Taxes, geography, and online-offline competition. NBER Working Paper No. W12242, Massachusetts Institute of Technology (MIT).

Forman, C., A. Ghose, A. Goldfarb. 2009. Competition between local and electronic markets: How the benefit of buying online depends on where you live, *Management Science*, **55**(1), 47-57

Glaeser, E.L., J. Kolko, A. Saiz. 2001. Consumer city. *Journal of Economic Geography* **1**(1) 27-50.

Goolsbee, A. 2000. In a world without borders: The impact of taxes on Internet commerce. *Quarterly Journal of Economics* **115**(2) 561-576.

Goolsbee, A. 2001. Competition in the computer industry: Online versus retail. *Journal of Industrial Economics* **49**(4) 487-499.

Greco, S. 2007. A world without bestsellers. *Inc. Magazine*, Harlan, IA (September).

Hausman, J., B.H. Hall, Z. Griliches. 1984. Econometric models for count data with an application to the patents-R & D relationship. *Econometrica* **52**(4) 909-938.

Jank, W., P.K. Kannan. 2005. Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Science* **24**(4) 623-634.

Mullahy, J. 1997. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics* **79**(4) 586-593.

Newey, W.K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* **36**(3) 231-250.

Prince, J. 2007. The beginning of online/retail competition and its origins: An application to personal computers. *International Journal of Industrial Organization* **25**(1) 139-156.

Rantisi, N.M. 2002. The competitive foundations of localized learning and innovation: The case of women's garment production in new york city. *Economic Geography* **78**(4) 441-462.

Raustiala K., C. Sprigman. 2006. The piracy paradox: Innovation and intellectual property in fashion design. *Virginia Law Review* **92**(8) 1687-1777.

Rubin, M.E. 2006. Holiday e-commerce spending up 26 percent with surge continuing the week before christmas, according to comscore networks. *PR Newswire*, Reston, Va. (December 28).

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

Tirole, J. 1988. *The theory of industrial organization*. The MIT Press, Cambridge, MA.

Varian, H. 2007. Why that hoodie your son wears isn't trademarked. *New York Times*, April 5.

Online Appendix

Battle of the Retail Channels:
How Product Selection and Geography Drive Cross-channel Competition

June 2009

Erik Brynjolfsson
MIT Sloan School of Management, erikb@mit.edu

Yu (Jeffrey) Hu
Krannert School of Management, Purdue University, yuhu@purdue.edu

Mohammad S. Rahman
Haskayne School of Business, University of Calgary, rahman@ucalgary.ca

Appendix A: Additional Statistics

Table A1: Descriptive Statistics of All Independent Variables

	Mean	Std. Deviation	Min	Max
NumStores	29.76	96.37	0	2715
Recency	429.57	295.65	1	958
Frequency	7.41	8.94	0	537
Monetary Value	23.40	11.60	0	179
Median Income	51,708.01	17754.77	3750	200,001
Population Density	0.24	0.64	0	15.39
Median Age of Female	37.52	4.33	14.3	85.8
Percentage with Bachelor's Degree	0.26	0.15	0	1
Percentage Female Population	0.51	0.02	0.01	0.82

Table A2: Economic Interpretation of the Impact of *NumStores* on Internet Demand for Popular Products

	Number of Stores	%Difference in Demand Compared to a Customer with Zero Store
Mean	30	-6.9
Median	7	-4.2
75^{th} Percentile	25	-6.5
95^{th} Percentile	110	-9.4

Appendix B: Extended Tables

This appendix presents the extended versions of the tables presented in the main body of the paper. In addition to the information presented in the original tables, each table includes an additional goodness-of-fit statistic – chi-squared statistic for the likelihood ratio (LR) test. In addition, each table reports the marginal effect of the number of stores (*NumStores*). Tables B1-B4 presents estimates from the probit model in addition to the estimates from the negative binomial model.

Table B1: Effect of Local Market Structure on Internet Demand
(This is an extended version of Table 2 presented in the main body of the paper)

	Probit (1)	Negative Binomial (2)
NumStores	-0.007*	-0.012*
	(0.003)	(0.006)
Recency	-0.152**	-0.254**
	(0.005)	(0.007)
Frequency	0.189**	0.375**
	(0.006)	(0.011)
Monetary Value	-0.063**	-0.106**
	(0.006)	(0.011)
Median Income	0.135**	0.253**
	(0.019)	(0.034)
Population Density	-0.020*	-0.039*
	(0.008)	(0.016)
Median Age of Female	-0.003**	-0.006**
	(0.001)	(0.002)
Percentage with Bachelor's Degree	0.160**	0.296**
	(0.045)	(0.081)
Percentage Female Population	-0.338	-0.708
	(0.230)	(0.414)
Intercept	-1.792**	-2.010**
	(0.243)	(0.440)
Log Likelihood	-51666.37	-91863.45
Likelihood Ratio Test (χ^2)	3486.60	3760.97
Sample Size	163,891	163,891
Marginal Effect of NumStores	-0.013	-0.012

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

Table B2: Effect of Local Market Structure on Different Products in the Internet Channel
(This is an extended version of columns (2) and (3) of Table 3 presented in the main body of the paper)

	Popular		Niche	
	Probit	Negative Binomial	Probit	Negative Binomial
	(1)	(2)	(3)	(4)
NumStores	-0.011**	-0.020**	0.002	0.006
	(0.003)	(0.006)	(0.004)	(0.009)
Recency	-0.138**	-0.234**	-0.175**	-0.327**
	(0.005)	(0.008)	(0.006)	(0.010)
Frequency	0.195**	0.387**	0.258**	0.566**
	(0.006)	(0.011)	(0.008)	(0.015)
Monetary Value	-0.062**	-0.102**	-0.124**	-0.247**
	(0.006)	(0.012)	(0.008)	(0.017)
Median Income	0.141**	0.268**	0.118**	0.237**
	(0.020)	(0.036)	(0.024)	(0.051)
Population Density	-0.017*	-0.036*	0.007	0.013
	(0.008)	(0.017)	(0.010)	(0.019)
Median Age of Female	-0.002*	-0.005*	-0.003**	-0.008**
	(0.001)	(0.002)	(0.001)	(0.003)
Percentage with Bachelor's Degree	0.183**	0.346**	0.049	0.120
	(0.047)	(0.085)	(0.058)	(0.120)
Percentage Female Population	-0.424	-0.820	-0.600*	-1.329*
	(0.237)	(0.434)	(0.289)	(0.588)
Intercept	-1.976**	-2.577**	-1.706**	-2.464**
	(0.251)	(0.463)	(0.310)	(0.640)
Log Likelihood	-48177.81	-81328.36	-28857.94	-40581.15
Likelihood Ratio Test (χ^2)	3147.34	3339.24	3157.49	3270.58
Sample Size	163,891	163,891	163,891	163,891
Marginal Effect of NumStores	-0.020	-0.020	0.005	0.006

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

Table B3: Effect of Local Market Structure on Overall Catalog Demand
(This is an extended version of column (4) of Table 3 presented in the main body of the paper)

	Probit (3)	Negative Binomial (4)
NumStores	-0.014**	-0.024**
	(0.003)	(0.005)
Recency	-0.152**	-0.242**
	(0.004)	(0.007)
Frequency	0.111**	0.225**
	(0.006)	(0.009)
Monetary Value	-0.042**	-0.070**
	(0.005)	(0.009)
Median Income	0.038*	0.074*
	(0.018)	(0.031)
Population Density	0.013	0.022
	(0.007)	(0.012)
Median Age of Female	0.007**	0.013**
	(0.001)	(0.002)
Percentage with Bachelor's Degree	-0.132**	-0.253**
	(0.043)	(0.075)
Percentage Female Population	0.227	0.416
	(0.217)	(0.374)
Intercept	-1.090**	-0.937*
	(0.228)	(0.397)
Log Likelihood	-61499.45	-109077.75
Likelihood Ratio Test (χ^2)	2533.74	2704.91
Sample Size	163,891	163,891
Marginal Effect of NumStores	-0.023	-0.024

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

Table B4: Effect of Local Market Structure on Different Products in the Catalog Channel
(This is an extended version of columns (5) and (6) of Table 3 presented in the main body of the paper)

	Popular		Niche	
	Probit	Negative Binomial	Probit	Negative Binomial
	(1)	(2)	(3)	(4)
NumStores	-0.018**	-0.032**	0.003	0.010
	(0.003)	(0.006)	(0.004)	(0.009)
Recency	-0.130**	-0.213**	-0.176**	-0.341**
	(0.005)	(0.007)	(0.006)	(0.011)
Frequency	0.116**	0.232**	0.174**	0.399**
	(0.006)	(0.010)	(0.008)	(0.016)
Monetary Value	-0.040**	-0.069**	-0.090**	-0.179**
	(0.005)	(0.010)	(0.007)	(0.017)
Median Income	0.052**	0.104**	-0.029	-0.071
	(0.018)	(0.032)	(0.024)	(0.051)
Population Density	0.010	0.016	0.026**	0.049**
	(0.007)	(0.013)	(0.009)	(0.018)
Median Age of Female	0.008**	0.014**	0.002	0.005
	(0.001)	(0.002)	(0.001)	(0.003)
Percentage with Bachelor's Degree	-0.128**	-0.251**	-0.127*	-0.292*
	(0.044)	(0.079)	(0.059)	(0.126)
Percentage Female Population	0.278	0.507	0.117	0.343
	(0.223)	(0.394)	(0.300)	(0.626)
Intercept	-1.478**	-1.692**	-0.598	-0.461
	(0.234)	(0.419)	(0.313)	(0.659)
Log Likelihood	-58117.12	-99787.82	-28625.89	-37217.06
Likelihood Ratio Test (χ^2)	2070.94	2203.25	2167.30	2253.70
Sample Size	163,891	163,891	163,891	163,891
Marginal Effect of NumStores	-0.031	-0.032	0.008	0.010

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

Appendix C: Results that Address Endogeneity

We address the endogeneity concern in both the binary response model and the count data model. In the probit model that is nonlinear, ordinary two stage least squares is not appropriate, hence we undertake the two-step approach that is due to Rivers and Vuong (1988). Following this two-step procedure for testing endogeneity, we first regress *NumStores* onto control variables and the instrumental variable – the number of local stores in 1994 (*NumStores94*). In particular, we estimate the following ordinary least squares (OLS) model and calculate the residuals:

$$NumStores = \beta_0 + \beta_1 z_1 + \beta_2 NumStores94 + v, \tag{C1}$$

where z_1 is a matrix of all independent variables except *NumStores*.

Next, we include the estimated residuals (i.e., $\hat{v}$) from this regression as an independent variable in our original probit model. A statistically significant coefficient of the residuals would signal *NumStores* is endogeneous, and a statistically insignificant coefficient of the residuals would signal *NumStores* is not endogeneous (Wooldridge 2002). Column (1) of Table C1 present the estimates from the second step.

The coefficient of the variable *Residuals* is not significantly different from zero. Thus, we fail to reject the null hypothesis that *NumStores* is exogenous. To further check the robustness of the endogeneity test, we estimate the probit model by using the instrumental variable for *NumStores*. More details regarding efficiently estimating probit models using instrumental variables can be found in Newey (1987). The estimates from the instrumented probit model are presented in column (2) of Table C1. Not surprisingly, the chi-square value associated with the Wald test of endogeneity is very small (0.02), and, once again, fails to reject the null hypothesis that *NumStores* is exogenous.

Table C1 Results that Address Endogeneity: Overall Demand in the Internet Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	-0.007	-0.007	-0.013	-0.013
	(0.004)	(0.004)	(0.008)	(0.008)
Recency	-0.152**	-0.152**	-0.254**	-0.254**
	(0.005)	(0.005)	(0.008)	(0.007)
Frequency	0.189**	0.189**	0.375**	0.375**
	(0.006)	(0.006)	(0.011)	(0.011)
Monetary Value	-0.063**	-0.063**	-0.107**	-0.107**
	(0.006)	(0.006)	(0.011)	(0.012)
Median Income	0.131**	0.131**	0.245**	0.245**
	(0.019)	(0.019)	(0.035)	(0.035)
Population Density	-0.020*	-0.020*	-0.037*	-0.038*
	(0.009)	(0.009)	(0.017)	(0.019)
Median Age of Female	-0.003**	-0.003**	-0.005**	-0.005**
	(0.001)	(0.001)	(0.002)	(0.002)
Percentage with Bachelor's Degree	0.161**	0.161**	0.302**	0.302**
	(0.047)	(0.047)	(0.084)	(0.085)
Percentage Female Population	-0.332	-0.332	-0.687	-0.689
	(0.234)	(0.234)	(0.421)	(0.429)
Residuals	0.001		0.004	
	(0.007)		(0.012)	
Intercept	-1.755**	-1.755**	-1.947**	-1.945**
	(0.245)	(0.245)	(0.443)	(0.449)
Log Likelihood	-50944.25		-90517.18	-90517.65
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 3387.10, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

To address the endogeneity concern in the count data model, we follow the two-step approach that is due to Wooldridge (1997). We include the *Residuals* calculated from OLS estimates of equation (C1) as an independent variable in the original negative binomial model. The estimates are presented in column (3) of Table C1. Consistent with our findings from probit models, we find that the coefficient of the variable *Residuals* is not significantly different from zero. Consequently, we fail to reject the null hypothesis that *NumStores* is exogenous. Nonetheless, we estimate the negative binomial model with the instrumental variable. We follow the two-stage approach described in Mullahy (1997). In the first stage, we calculate the predicted *NumStores* from OLS estimates of equation (C1), and then include the predicted value of *NumStores* when estimating the negative binomial model in the second stage. The estimates are presented in column (4) of Table C1. As expected, the Hausman specification test for endogeneity using the coefficients in column (4) of Table C1 produces a t statistic of -0.194. This suggests that we fail to reject the null hypothesis that *NumStores* is not endogenous.

Wooldridge (2002) mentions that the two-step approach that we have followed for testing endogeneity in both probit model and negative binomial model is a very robust approach. Not surprisingly, we reach the same conclusion whether we use the two-step approach or we do Hausman specification tests using the instrumental variable results: we fail to find evidence that *NumStores* is endogenous. Reassuringly, given that we have a large sample, these two approaches produce similar coefficient estimates. In addition, coefficient estimates in columns (1) and (2) of Table 2 are similar to coefficient estimates in Table C1. This provides further evidence that our results are robust to using estimators that address the endogeneity concern. Please note that Tables C2-6 reports results that address endogeneity concerns for rest of the models estimated in this paper. Consistently we fail to find evidence that *NumStores* is endogenous.

Table C2 Results that Address Endogeneity: Demand for Popular Products in the Internet Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	-0.013**	-0.013**	-0.024**	-0.024**
	(0.004)	(0.004)	(0.008)	(0.008)
Recency	-0.138**	-0.138**	-0.234**	-0.234**
	(0.005)	(0.005)	(0.008)	(0.008)
Frequency	0.194**	0.194**	0.386**	0.386**
	(0.006)	(0.006)	(0.011)	(0.011)
Monetary Value	-0.062**	-0.062**	-0.103**	-0.103**
	(0.006)	(0.006)	(0.012)	(0.012)
Median Income	0.137**	0.137**	0.261**	0.261**
	(0.020)	(0.020)	(0.037)	(0.037)
Population Density	-0.016	-0.016	-0.031	-0.032
	(0.009)	(0.009)	(0.018)	(0.020)
Median Age of Female	-0.002*	-0.002*	-0.004*	-0.004*
	(0.001)	(0.001)	(0.002)	(0.002)
Percentage with Bachelor's Degree	0.187**	0.187**	0.358**	0.358**
	(0.048)	(0.048)	(0.089)	(0.089)
Percentage Female Population	-0.394	-0.394	-0.751	-0.754
	(0.241)	(0.241)	(0.443)	(0.449)
Residuals	0.004		0.011	
	(0.007)		(0.013)	
Intercept	-1.955**	-1.955**	-2.544**	-2.540**
	(0.252)	(0.252)	(0.467)	(0.471)
Log Likelihood	-47492.23		-80109.99	-80110.91
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 3049.90, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

Table C3 Results that Address Endogeneity: Demand for Niche Products in the Internet Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	0.004	0.004	0.007	0.007
	(0.006)	(0.006)	(0.011)	(0.011)
Recency	-0.175**	-0.175**	-0.327**	-0.327**
	(0.006)	(0.006)	(0.010)	(0.010)
Frequency	0.259**	0.259**	0.568**	0.568**
	(0.008)	(0.008)	(0.015)	(0.016)
Monetary Value	-0.123**	-0.123**	-0.245**	-0.245**
	(0.008)	(0.008)	(0.018)	(0.018)
Median Income	0.113**	0.113**	0.227**	0.227**
	(0.025)	(0.025)	(0.051)	(0.052)
Population Density	0.005	0.005	0.011	0.010
	(0.010)	(0.010)	(0.021)	(0.021)
Median Age of Female	-0.003*	-0.003*	-0.008**	-0.008**
	(0.001)	(0.001)	(0.003)	(0.003)
Percentage with Bachelor's Degree	0.054	0.054	0.137	0.137
	(0.060)	(0.060)	(0.124)	(0.125)
Percentage Female Population	-0.640*	-0.640*	-1.401*	-1.401*
	(0.294)	(0.294)	(0.596)	(0.607)
Residuals	-0.003		-0.002	
	(0.009)		(0.018)	
Intercept	-1.643**	-1.643**	-2.344**	-2.343**
	(0.311)	(0.311)	(0.643)	(0.654)
Log Likelihood	-28450.27		-40014.87	-40014.94
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, $p<0.01$
* Significantly different from zero, $p<0.05$

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 3106.54, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

Table C4 Results that Address Endogeneity: Overall Demand in the Catalog Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	-0.012**	-0.012**	-0.019**	-0.019**
	(0.004)	(0.004)	(0.007)	(0.007)
Recency	-0.153**	-0.153**	-0.243**	-0.243**
	(0.004)	(0.004)	(0.007)	(0.007)
Frequency	0.110**	0.110**	0.224**	0.224**
	(0.006)	(0.006)	(0.010)	(0.010)
Monetary Value	-0.042**	-0.042**	-0.070**	-0.070**
	(0.005)	(0.005)	(0.009)	(0.010)
Median Income	0.037*	0.037*	0.073*	0.073*
	(0.018)	(0.018)	(0.031)	(0.032)
Population Density	0.010	0.010	0.017	0.017
	(0.008)	(0.008)	(0.013)	(0.014)
Median Age of Female	0.007**	0.007**	0.013**	0.013**
	(0.001)	(0.001)	(0.002)	(0.002)
Percentage with Bachelor's Degree	-0.132**	-0.132**	-0.254**	-0.254**
	(0.044)	(0.044)	(0.077)	(0.078)
Percentage Female Population	0.170	0.170	0.313	0.305
	(0.221)	(0.221)	(0.379)	(0.371)
Residuals	-0.007		-0.012	
	(0.006)		(0.011)	
Intercept	-1.055**	-1.055**	-0.872*	-0.867*
	(0.229)	(0.229)	(0.399)	(0.402)
Log Likelihood	-60788.13		-107816.56	-107824.08
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 2491.35, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

Table C5 Results that Address Endogeneity: Demand for Popular Products in the Catalog Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	-0.016**	-0.016**	-0.028**	-0.028**
	(0.004)	(0.004)	(0.007)	(0.007)
Recency	-0.130**	-0.130**	-0.213**	-0.214**
	(0.005)	(0.005)	(0.007)	(0.007)
Frequency	0.115**	0.115**	0.231**	0.231**
	(0.006)	(0.006)	(0.010)	(0.010)
Monetary Value	-0.041**	-0.041**	-0.070**	-0.070**
	(0.005)	(0.005)	(0.010)	(0.010)
Median Income	0.052**	0.052**	0.104**	0.104**
	(0.018)	(0.018)	(0.033)	(0.033)
Population Density	0.007	0.007	0.011	0.010
	(0.008)	(0.008)	(0.014)	(0.016)
Median Age of Female	0.008**	0.008**	0.014**	0.014**
	(0.001)	(0.001)	(0.002)	(0.002)
Percentage with Bachelor's Degree	-0.129**	-0.129**	-0.252**	-0.252**
	(0.045)	(0.045)	(0.081)	(0.082)
Percentage Female Population	0.221	0.221	0.407	0.394
	(0.226)	(0.226)	(0.400)	(0.392)
Residuals	-0.007		-0.012	
	(0.006)		(0.011)	
Intercept	-1.447**	-1.447**	-1.635**	-1.624**
	(0.236)	(0.236)	(0.421)	(0.424)
Log Likelihood	-57442.35		-98624.36	-98634.97
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, $p<0.01$
* Significantly different from zero, $p<0.05$

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 2019.84, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

Table C6 Results that Address Endogeneity: Demand for Niche Products in the Catalog Channel

	Two-step Probit (1)	IV Probit[a] (2)	Two-step Negative Binomial (3)	IV Negative Binomial[b] (4)
NumStores	0.008	0.008	0.022	0.022
	(0.006)	(0.006)	(0.012)	(0.012)
Recency	-0.177**	-0.177**	-0.343**	-0.343**
	(0.006)	(0.006)	(0.011)	(0.010)
Frequency	0.173**	0.173**	0.397**	0.397**
	(0.008)	(0.008)	(0.016)	(0.017)
Monetary Value	-0.089**	-0.089**	-0.177**	-0.177**
	(0.008)	(0.008)	(0.017)	(0.018)
Median Income	-0.032	-0.032	-0.075	-0.075
	(0.024)	(0.024)	(0.052)	(0.053)
Population Density	0.021*	0.021*	0.037	0.037
	(0.010)	(0.010)	(0.019)	(0.019)
Median Age of Female	0.003	0.003	0.005	0.005
	(0.001)	(0.001)	(0.003)	(0.003)
Percentage with Bachelor's Degree	-0.140*	-0.140*	-0.322*	-0.322*
	(0.061)	(0.061)	(0.130)	(0.131)
Percentage Female Population	0.029	0.029	0.147	0.148
	(0.304)	(0.304)	(0.632)	(0.596)
Residuals	-0.012		-0.028	
	(0.009)		(0.018)	
Intercept	-0.534	-0.534	-0.322	-0.324
	(0.314)	(0.314)	(0.661)	(0.659)
Log Likelihood	-28294.35		-36799.56	-36799.67
Sample Size	161,856	161,856	161,856	161,856

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

[a] The estimates are obtained using Newey's (1987) minimum chi-squared estimator. The chi-square model fit statistics is 2187.85, suggesting the model is significant.

[b] Robust standard errors are listed in parentheses. The standard errors obtained from bootstrapping are almost identical to the standard errors reported here.

Appendix D: Robustness Check (Including Wal-Mart and Target Stores)

This appendix presents robustness of our results, even after recoding the variable *NumStores* (including Wal-Mart and Target stores).

Table D1: Effect of Local Market Structure (including Wal-Mart and Target) on Internet Demand

	Internet Channel		
	Overall	Popular	Niche
NumStores	-0.009	-0.016*	0.007
	(0.006)	(0.007)	(0.009)
Recency	-0.254**	-0.234**	-0.327**
	(0.007)	(0.008)	(0.010)
Frequency	0.375**	0.387**	0.566**
	(0.011)	(0.011)	(0.015)
Monetary Value	-0.106**	-0.102**	-0.247**
	(0.011)	(0.012)	(0.017)
Median Income	0.253**	0.268**	0.237**
	(0.034)	(0.036)	(0.051)
Population Density	-0.043**	-0.041*	0.012
	(0.016)	(0.017)	(0.019)
Median Age of Female	-0.006**	-0.005*	-0.008**
	(0.002)	(0.002)	(0.003)
Percentage with Bachelor's Degree	0.284**	0.332**	0.117
	(0.081)	(0.086)	(0.120)
Percentage Female Population	-0.745	-0.855*	-1.344*
	(0.415)	(0.435)	(0.590)
Intercept	-1.992**	-2.556**	-2.461**
	(0.439)	(0.463)	(0.640)
Log Likelihood	-91864.51	-81330.29	-40581.09
Sample Size	163,891	163,891	163,891

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05

Table D2: Effect of Local Market Structure (including Wal-Mart and Target) on Catalog Demand

	Catalog Channel		
	Overall	Popular	Niche
NumStores	-0.027**	-0.036**	0.007
	(0.005)	(0.006)	(0.009)
Recency	-0.242**	-0.213**	-0.341**
	(0.007)	(0.007)	(0.011)
Frequency	0.225**	0.233**	0.399**
	(0.009)	(0.010)	(0.016)
Monetary Value	-0.070**	-0.069**	-0.179**
	(0.009)	(0.010)	(0.017)
Median Income	0.074*	0.105**	-0.070
	(0.031)	(0.032)	(0.051)
Population Density	0.024*	0.018	0.052**
	(0.012)	(0.013)	(0.018)
Median Age of Female	0.013**	0.014**	0.005
	(0.002)	(0.002)	(0.003)
Percentage with Bachelor's Degree	-0.240**	-0.236**	-0.278*
	(0.075)	(0.079)	(0.126)
Percentage Female Population	0.484	0.592	0.391
	(0.376)	(0.397)	(0.629)
Intercept	-0.954*	-1.712**	-0.480
	(0.397)	(0.419)	(0.659)
Log Likelihood	-109075.49	-99784.81	-37217.51
Sample Size	163,891	163,891	163,891

Standard errors are listed in parentheses.
** Significantly different from zero, p<0.01
* Significantly different from zero, p<0.05