

DRAFT: Comments welcome. Please do not quote without permission.

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

May 2007

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

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

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 almost entirely neglected this fundamental dimension of competition. 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 a data set on the local market structure, and then match this data set on local market structure to a data set on consumer demand that is through direct channels that include Internet and catalog channels. Our analyses provide strong evidence that the local market structure can significantly explain the variation in demand through direct channels, even after controlling for the relevant demographic and socioeconomic variables in each local market. We find that the impact of the local market structure on consumer demand through direct channels is smaller in size for niche products than that for popular products. In addition, we identify the role of demand for popular products and demand for niche products in shaping the impact of local stores on the catalog channel and the Internet channel. The sales of niche products, which are often unavailable in physical stores, are largely immune from competition by traditional retailers. Since the Internet channel sells proportionately more niche products than the catalog channel, the level of competition between the Internet channel and local stores is lower than the level of competition between the catalog channel and local stores.

Keywords: Internet, competition, retailing, channel, niche products, geography.

*The authors are listed in alphabetical order.

Acknowledgements: We thank Jason Abrevaya, Prabuddha De, Alok Gupta, Vallabh Sambamurthy, seminar participants at Purdue University, and participants at 2007 Big Ten Information Systems Research Symposium for valuable comments on this paper. We thank Zhulei Tang and Krannert Computing Center for providing computing resources. Purdue Research Foundation and MIT Center for Digital Business provided generous funding.

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 competition with traditional brick-and-mortar retailers. In almost every product category sold on the Internet, consumers have the option of buying from a traditional retailer instead. Although traditional retailers vastly outsell Internet retailers in most of these categories, research on Internet retailing has almost entirely neglected this fundamental dimension of competition.¹ How and where can Internet retailers win this battle?

In order to compete with brick-and-mortar stores, Internet retailers have made large investments in delivery infrastructures, Web technologies, marketing promotions, and customer services. For instance, Jeff Wilke, SVP of Amazon, identified the “instant gratification” available in traditional stores as the key challenge to Amazon’s growth.² As a result, in order to compete with brick-and-mortar retailers, many Internet firms have built delivery centers all across the U.S., speeding up the delivery of their products to consumers. In order to attract consumers, who usually do not pay shipping and handling charges when purchasing from local stores, Internet retailers frequently offer free-shipping discounts. Finally, Internet retailers have invested heavily in technologies that allow consumers to carefully inspect and sample products before making purchases, and they offer customer services that are as good as those offered by brick-and-mortar retailers, alleviating consumers’ concern about returns and refunds.

Despite the widely-accepted notion that Internet commerce competes with brick-and-mortar commerce, and despite the millions of dollars that have been invested in the competition between Internet retailers and brick-and-mortar retailers, our understanding of this type of competition is remarkably limited. In

¹ ComScore, an online market research company, estimates that in 2006 non-travel Internet retail has accounted for approximately 7 percent of U.S. consumers’ retail spending excluding gas, autos and food (Rubin 2006).

particular, there is little academic research that studies whether the level of this competition varies in different local markets and for different products. This paper attempts to address these questions using a unique combination of data sets.

There exists strong theoretical support for that the competition between Internet retailing and brick-and-mortar retailing varies in different local markets. In markets where consumers have access to many brick-and-mortar stores, the competition between Internet retailing and brick-and-mortar retailing is intensified; while in markets where consumers are under-served by local stores, Internet retailing faces little or no competition. Thus, it is likely that the local market structure, i.e., the number of local stores within the distance customers are willing to travel for purchasing consumer products, will affect consumer demand through the Internet channel. However, empirical research on this topic is lacking.

The lack of research on this topic is at least partly due to the lack of data on local market structure: until recently, it has been difficult for researchers to obtain such data. Only recent technological advances have made this data available. In this paper, we collect a unique data set on the local market structure. We then match this data set on local market structure to a data set on consumer demand through direct channels that include the Internet and catalog channels. Our analyses provide strong evidence that the local market structure can significantly explain the variation in demand through direct channels, even after controlling for the relevant demographic and socioeconomic variables in each local market.

In theory, the competition between Internet retailing and brick-and-mortar retailing should also vary for different products. When an Internet retailer sells popular products that are likely to be widely available in local stores, the competition between such an Internet retailer and local stores is intensified. However, as revealed by recent research on the “Long Tail”, Internet retailers may face little or no competition if they sell niche products that are unlikely to be available in local stores (Brynjolfsson et al. 2003; Anderson 2004; Brynjolfsson et al. 2006). Reassuringly, our empirical results are consistent with this

² The authors’ interview at Amazon Headquarters, Seattle, Washington, February, 2005.

theoretical prediction. We find that the impact of the local market structure on consumer demand through direct channels is smaller in size for niche products than that for popular products.

Although, in many respects, the Internet channel is similar to 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 confirms that finding of previous research. An interesting implication of that finding is the relative competition among the Internet, catalog and local channels. Specifically, since the Internet channel sells proportionately more niche products than the catalog channel, it is likely that the level of competition between the Internet channel and local stores is lower than the level of competition between the catalog channel and local stores. We find empirical results that are consistent with this prediction.

Understanding how the level of the competition between direct channels and local stores varies in different local markets and for different products has both important managerial implications and economic consequences. Answers to this question would help economists and public-policy makers more accurately estimate the new value created by Internet commerce, guiding the government in its policies toward the Internet commerce. In addition, our study shows that the local market structure variable can have a significant impact on consumer demand and this impact differs for different products.

Historically, firms in the direct retailing industry have long used various measures to segment consumers and treated different consumer segments with different marketing strategies. Thus, our results suggest that including the local market structure variable in these firms' marketing decisions can make their marketing strategies more effective.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the relevant literature on this topic. In Section 3, we discuss our data collection methodology and data sources. We present our empirical analyses and results in Section 4. We discuss the implications of our study in Section 5. Finally, we conclude the paper in Section 6.

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 strongly competes with local stores. As the number of brick-and-mortar stores increases, the competition becomes more intense and the profit of both the direct retailer and brick-and-mortar stores decreases.

In a landmark paper, Bresnahan and Reiss (1991) find that the number of firms in a local market is correlated with market size variables. They then show that as the number of competing firms in a local market increases, the competition becomes more intensified and firms' profit margins fall. More recently, Campbell and Hopenhayn (2005) find that the competition is tougher in larger markets. These findings support the hypothesis that a direct retailer would be in an advantageous position in local markets where there are no or few brick-and-mortar stores, compared with in local markets where many local stores exist. Holding prices constant, a direct retailer's demand will fall as the number of local stores increases.

Previous research suggests that greater selections, lower prices, and convenience are primary drivers in enticing customers to the catalog channel (Bitran and Mondschein 1996). Internet markets can improve consumer welfare through wider product selection (Brynjolfsson et al. 2003). Consumer demand through the Internet will be higher in local markets where local prices and sales tax rates are high (Goolsbee 2000, 2001, Chiou 2005, Prince 2006, Ellison and Ellison 2006). However, the impact of the local market structure, i.e., the number of local stores, on consumer demand through the direct channels that include the Internet and catalog channels 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 geographical variables can have an impact on consumer behavior online-- Sinai and Waldfogel (2004), Forman et al. (2006), although it also differs from them in many aspects. Sinai and Waldfogel (2004) find that an individual is more likely to

connect to the Internet if there are more content that is of interest to her on the Internet and there are less content that is of interest to her locally. Their results suggest that the Internet medium competes with offline media, but they do not study the competition between Internet retailing and brick-and-mortar retailing, which is the focus of our paper. Forman et al. (2006) find evidence that the existence of a discount store or a large bookstore in a geographical location decreases the likelihood that a popular book will appear in Amazon's list of top 10 bestselling books for that geographical location. Our paper differs significantly from their paper by studying the effect of the number of local stores on individual-level demand and discussing how this effect varies for different channels and different products—rather than studying the effect of store existence on aggregate-level demand. More specifically, we directly measure the individual-level demand, while they use aggregate sales ranking for top products at the geographical location-level to make inferences about demand. Furthermore, we directly measure the number of physical stores at the zip code level, while they use the existence of stores at the level of less specific geographical locations which include large cities and small towns.

Recently there are a few papers in marketing that have leveraged spatial data to understand consumers' behavior (e.g., Bradlow et al. 2005 for a review of this emerging literature). 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 sold through a publisher's website. We contribute to this nascent literature by highlighting how the local market structure, which varies with geographic location, can affect a consumer's purchasing behavior at direct channels and how this effect varies across the Internet and catalog channels.

3. Data

For this study, we have collected data from several sources. Our customer demand data comes from a large retailer of women's clothing products.³ The retailer primarily operates in two channels: the catalog

³ The retailer requests to remain anonymous.

channel and the Internet channel, with both channels contributing almost equally to the firm's revenue.⁴ We have information regarding all the transactions made starting from May 19, 2003 to June 15, 2006 through the primary channels: the catalog channel (mail and telephone orders) and the Internet channel (website). For each item purchased from the retailer, we have information regarding the price paid, date of transaction, customer's unique id, whether or not the item was returned, channel (e.g., mail, telephone, or the web) used to purchase the item, and transaction id. Overall we have records of about 7 million transactions that were made by about 1 million unique customers. Moreover, we have information regarding what catalogs and emails each customer received between January 2005 and June 2006. We also know each customer'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 customer. In addition, the data allows us to determine the local market structure for each customer.

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. Also, 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 structure comes from two leading providers of store-directory services: Yahoo Local (<http://local.yahoo.com>), a leading online portal that provides information of local businesses, and Superpages.com (<http://www.superpages.com>), a Verizon spin-off that is a prominent provider of yellow pages and information services. We have first 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 written a set of web-crawling programs to collect data from each website. For each zip code,

⁴ The retailer also has a physical store. We do not have any information regarding the transactions made in the physical store. Note that the physical store accounts for a negligible amount of overall sales.

we have collected the number of women’s clothing stores listed on Superpages.com that are within 5 miles, 10 miles, 15 miles, 20 miles, 25 miles, and 50 miles from the center of that zip code. Similarly, we have collected the number of women’s clothing store listed on Yahoo Local that are within 5 miles, 10 miles, 15 miles, 25 miles, and 50 miles from the center of each zip code. In order to ensure consistency of the data collected, we have collected this data three times from each website between the last week of October 2006 and the end of November 2006. The differences among the data collected in these three snapshots are very small and negligible, and we have not found any systematic or significant differences among the data collected in these three snapshots. For this study, we will use data obtained from the last collection.

In order to alleviate any concerns regarding the accuracy of the local market structure data and any site-specific effect, we simultaneously collect this data from two independent sites and check whether the data collected from Superpages.com is similar to the data collected from Yahoo Local. We are able to find information for about 41,219 zip codes from both sites. For each common radius, we calculate the Pearson correlation coefficient between the number of stores listed on Superpages.com and the number of stores listed on Yahoo Local. We find these two sets of numbers are extremely highly correlated. The correlation coefficients are presented in Table 1. This gives us even more confidence in the reliability of our data on local market structure. Since Superpages.com is the largest provider of directory services, we present our results using the data collected from this site. As expected, the qualitative nature of the results in this paper does not change if we use the data collected from Yahoo Local instead.

Table 1: Pearson Correlation of the Number of Stores Listed

Radius	Correlation
5 miles	0.911**
10 miles	0.951**
15 miles	0.965**
25 miles	0.973**
50 miles	0.974**

** Significantly different from zero, $p < 0.01$

A customer's demographic and socioeconomic variables such as her income, age, education, and gender may influence that customer's demand. Since we observe each customer's home zip code, we are able to collect these demographic and socioeconomic variables at the zip code level from the U.S. Census 2000.⁵ These variables, along with variables that indicate whether each local market is in an urban area or not, will be used as control variables in our analyses.

4. Empirical Analyses

In our analyses we focus on Year 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 a customer's demand. The retailer promotes its products by sending catalogs and emails. Although every customer with valid email address receives all emails, every customer does not receive all catalogs within a time period. It is possible that a customer's demand within the time period from January 1, 2006 to June 15, 2006 is influenced by the number of catalogs received during this time period as well as the catalogs received before January 1, 2006. 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 customers who have received all the catalogs that were sent out between November 1, 2005 (61 days prior to January 1, 2006) and June 15, 2006.⁶ Since our analyses focus on transactions that occurred between January 1, 2006 and June 15, 2006, we will later use the earlier transaction data to calculate historical purchasing measures and use them as controls for customer heterogeneity. We have found 183,023 customers each of whom has received all the catalogs sent out during the period between November 1, 2005 and June 15, 2006. There are a number of customers who are from outside the 50 U.S. states and

⁵ The demographic and socioeconomic variables used in this paper are contained in the Summary File 1 and Summary File 3 of U.S. Census 2000.

⁶ In total, each customer in our sample received 13 catalogs between November 1, 2005 and June 15, 2006. Catalogs were sent out on the following dates: Nov 21, Dec 20, Jan 2, Jan 16, Feb 6, Feb 20, Mar 6, Mar 27, Apr 7, Apr 24, May 08, and May 24, June 5.

the District of Columbia; we have excluded those customers from our final sample. Also, since the retailer has only one physical store in Florida, it needs to collect sales tax on sales to Florida customers. Although our results remain practically unchanged even if we include Florida customers in our analyses, we have excluded all the Florida customers in order to eliminate the difference in sales taxes as a confounding factor. Correspondingly, we have retained 163,933 customers for our analyses. Table 2 presents the descriptive statistics for our sample. During the period between January 1, 2006 and June 15, 2006, the average number of items bought from the retailer is 0.94 for the customers in our sample, with approximately 47% of the customers' purchases occurring through the Internet channel.

Table 2: Descriptive Statistics on Customers' Demand

	Average Number of Items Purchased
Overall	0.94
Internet Channel	0.44
Catalog Channel	0.50
Sample Size	163,933

4.1. Initial Results

First, we study whether the local market structure can have an impact on the probability of purchasing from the direct retailer. Assuming y_i is an indicator of whether customer i had positive demand between January 1, 2006 and June 15, 2006, and X_i is a vector of explanatory variables, we can estimate the effect of the local market structure on y_i using the following Probit model:

$$P(y_i = 1 | X_i) = \Phi(X_i\beta). \tag{1}$$

We let variable *NumStores* be the natural log of the number of local stores listed on Superpages.com within 5 miles of each customer's home zip code.⁷ In column (1) of Table 3, we present the estimation results of such a Probit model with the explanatory variables being *NumStores* and an intercept.

Reassuringly, this result and all the results that follow remain qualitatively unchanged, even if we use the

data collected from Yahoo Local instead of the data collected from Superpages.com, or if we use the number of local stores within 10 miles (or other distances) instead of 5 miles.⁸ In addition, although we only report the results using a Probit model, our findings are robust to using a Logit model as an alternative.

Our coefficient of interest, the coefficient on variable *NumStores*, is negative and significantly different from zero. This indicates that customers with more traditional stores nearby are less likely to purchase anything from the direct retailer, and this effect is economically and statistically significant. Direct retailers that operate Internet and catalog channels do directly compete with brick-and-mortar retailers.

Next, we investigate the impact of the local market structure on the overall demand at the direct retailer. Note that the number of items purchased, which is count data, can be assumed to follow a Poisson distribution. However, a Poisson regression model assumes that the mean and the variance are equal. If this assumption is not satisfied, a Poisson regression would provide consistent parameter estimates, but the standard errors would be underestimated. In our data, the variance exceeds the mean which causes over-dispersion. Consequently, 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):

$$f(y_i | X_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0,1,2,3,\dots \quad (2)$$

where: y_i is the number of items purchased by customer i between January 1, 2006 and June 15, 2006; X_i is a vector of explanatory variables; $E(y_i | X_i) = \mu_i = \exp(X_i\beta + \varepsilon_i)$ is the conditional

⁷ The natural log of zero is not defined. Thus we add one to the number of local stores before taking the natural log.

⁸ The number of local stores within 5 miles is highly correlated with the numbers of local stores within 10 miles and other distances.

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; Greene 2002).⁹

The estimates of the negative binomial regression model are presented in column (2) of Table 3. Once again, the coefficient associated with the variable *NumStores* is negative and significantly different from zero. This suggests that as the number of nearby local stores increases, a consumer purchases fewer items from the direct retailer. We interpret this result as evidence that the local market structure can have an impact on consumer demand at the direct retailer.

Table 3: The Effect of Local Market Structure on Demand at the Direct Retailer

	Probit (1)	Negative Binomial (2)
NumStores	-0.007** (0.002)	-0.010** (0.003)
Intercept	-0.790** (0.006)	-0.040** (0.010)
Log Likelihood	-84,430.56	-169,135.85
Sample Size	163,933	163,933

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

4.2. Controlling for Demographic and Socioeconomic Variables and Urban Area versus Non-urban Area Effect

A customer's demographic and socioeconomic variables such as her income, age, education, and gender can influence her demand. A concern regarding the results in Table 3 may be that the variable of the number of local stores embodies differences in these demographic and socioeconomic variables, and in turn, these variables' effect on the customer's demand. We will address this issue by doing the following. We will collect these demographic and socioeconomic variables from U.S. Census 2000 at the zip code

⁹ A specification test for rejecting the Poisson regression model can be carried out by testing the hypothesis $\theta = 0$.

level, and use these variables as control variables when we estimate Probit and negative binomial regression models. Thus, control variables in our analysis will include the median household income in natural log, percentage of female population, percentage of population with at least a Bachelor's degree, and median age of the female population, all at the level of the customer's home zip code.¹⁰

In addition, whether a customer lives in an urban area or a non-urban area may have an influence on that customer's demand (Glaeser et al. 2001). In order to address the issue of whether the variable of the number of local stores represents this urban area versus non-urban area effect, we will add a population density variable, defined as the population per square mile divided by 10,000, as a control variable.

Column (1) and column (2) of Table 4 present the results obtained from a Probit model and a negative binomial regression model, after controlling for demographic and socioeconomic variables, urban area versus non-urban area effect. The results in column (1) and column (2) of Table 4 indicate that income and age do influence customer demand, separately from the number of local stores. But other control variables are not statistically significant. More importantly, the effect of the number of local stores is still negative and significant. We interpret this as evidence that customers with more brick-and-mortar stores nearby have a lower demand at the direct retailer, and this competition effect of local stores on direct retailing persists even after controlling for demographic and socioeconomic variables and urban area versus non-urban area effect.¹¹ As expected, the fit is better for both models after including all these control variables. This is shown by the higher log likelihood numbers (smaller in absolute value) in column (1) and column (2) of Table 4 than in Table 3. Our likelihood ratio tests show that the basic specifications in column (1) and column (2) of Table 3 are rejected, therefore, we need to include these demographic and socioeconomic variables and urban area versus non-urban area effect.

¹⁰ 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.

¹¹ If some of these control variables are highly correlated with one another and with variable *NumStores*, then adding these control variables may cause multicollinearity, which leads to statistically insignificant coefficients. We calculate the correlation coefficients among control variables and variable *NumStores* and find none of these correlation coefficients is high enough to cause multicollinearity concerns.

Table 4: The Effect of Local Market Structure on Demand, Controlling for Socioeconomic Factors and Historical Purchase

	Probit (1)	Negative Binomial (2)	Probit (3)	Negative Binomial (4)	Probit (5)	Negative Binomial (6)
NumStores	-0.009** (0.003)	-0.014** (0.004)	-0.009** (0.002)	-0.013** (0.003)	-0.012** (0.003)	-0.019** (0.004)
Recency			-0.037** (0.007)	-0.067** (0.010)	-0.038** (0.007)	-0.069** (0.010)
Frequency			0.231** (0.005)	0.410** (0.008)	0.232** (0.005)	0.410** (0.008)
Monetary Value			-0.028** (0.005)	-0.053** (0.008)	-0.028** (0.005)	-0.053** (0.008)
Median Income	0.095** (0.015)	0.164** (0.024)			0.083** (0.016)	0.146** (0.024)
Population Density	0.003 (0.006)	0.003 (0.010)			0.000 (0.006)	-0.002 (0.010)
Median Age of Female	0.003** (0.001)	0.005** (0.001)			0.004** (0.001)	0.006** (0.001)
Percentage with Bachelor's Degree	-0.036 (0.037)	-0.067 (0.057)			0.013 (0.037)	0.019 (0.057)
Percentage Female Population	-0.160 (0.189)	-0.312 (0.288)			-0.041 (0.191)	-0.124 (0.287)
Intercept	-1.845** (0.197)	-1.837** (0.304)	-0.887** (0.047)	-0.227** (0.073)	-1.894** (0.203)	-1.941** (0.310)
Log Likelihood	-84375.8	-169060.9	-82868.0	-167171.6	-82809.8	-167090.7
Sample Size	163,891	163,891	163,933	163,933	163,891	163,891

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

If anything, this competition effect of local stores on direct retailing becomes larger in size after adding in these control variables, as evidenced by the fact that the coefficient on variable *NumStores* becomes more negative. This result is hardly surprising—without including some of these control variables such as income that are positively correlated with demand, the variable of the number of local stores can embody

these positive effects on demand and become less negative when these control variables are not controlled for. Once those variables are controlled for, the coefficient on variable *NumStores* reflects less of these positive effects on demand, thus becoming more negative.

4.3. Controlling for Historical Purchasing Measures

Historical purchasing measures are widely used in the direct retailing industry to segment customers into loyal and non-loyal customers and to control for customer heterogeneity. In this industry as well as in the academic literature, “Recency”, “Frequency”, and “Monetary Value” measures, also known as the “RFM” measures, have been widely used to measure customers’ historical purchasing behavior and segment them into different segments (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 customer paid over a time period. Next we will use these RFM measures, in natural log, of customers’ historical purchasing behavior in the time period prior to January 1, 2006 as control variables for customer heterogeneity.

Column (3) and column (4) of Table 4 present the results obtained from a Probit model and a negative binomial regression model, after controlling for historical purchasing measures. The results in column (3) and column (4) of Table 4 indicate that historical purchasing measures can explain a lot of the variation in demand in the current period, and the coefficients on “RFM” measures are highly significant with large t-statistics. This is consistent with what previous research has found. But more importantly for our research, the coefficient on variable *NumStores* is negative and highly significant, even after controlling for the “RFM” measures which have been widely used by direct retailers to segment consumers and to treat different consumer segments with different marketing strategies. Thus, our results suggest that including the local market structure variable in these firms’ marketing decisions can make their marketing strategies more effective.

To be complete, we control for demographic and socioeconomic variables, urban area versus rural area effect, as well as historical purchasing measures, in column (5) and column (6) of Table 4. The coefficient on variable *NumStores* remains negative and highly significant. In addition, results in Table 4 from column (1) through column (6) show that the sign and the size of the coefficient on variable *NumStores* do not change wildly when different specifications (control variables) of the model are used, suggesting the robustness of our result.

To further interpret the effect of the local market structure on customer demand at the direct retailer, we calculate how customers' demand will change when the number of local stores increases from the 25th percentile of its sample distribution to the 50th percentile, and then to the 75th percentile. The distribution of the number of local stores reaches its 25th percentile at 0, 50th percentile at 7, and 75th percentile at 30. We will use the coefficients reported in column (6) of Table 4 in our calculation. Holding all the control variables in our analysis constant, increasing the number of local stores from 0 store to 7 stores reduces customers' demand at the direct retailer by 4.0%. Everything else being equal, increasing the number of local stores from 7 stores to 30 stores reduces customers' demand at the direct retailer by 2.6%.

4.4. Robustness Check

An even stronger approach to test the robustness of our results is to estimate the effect of local market structure on consumer demand using dummy variables instead of a continuous measure of the number of stores. This approach also allows us to detect the existence of non-monotonicity if such an effect is indeed non-monotonic. We first create a dummy variable *StoreAbove0* indicating whether there is at least one store within 5 miles of the customer's home zip code. We then replace variable *NumStores* with this dummy variable *StoreAbove0* when we estimate our negative binomial regression model, while keeping all control variables intact. The coefficient on dummy variable *StoreAbove0* is reported in Table 5. We have omitted the coefficients on control variables for the sake of brevity—these coefficients are similar to the coefficients reported in column (6) of Table 4. We then create the second 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 of the customer’s home zip code, and the third dummy variable *StoreAbove30* indicating whether there are more than 30 stores (which is the 75th percentile of the distribution of the number of stores) within 5 miles of the customer’s home zip code. We replace variable *NumStores* with these two dummy variables, respectively, when we estimate a negative binomial regression model. These results are also reported in Table 5.

Table 5: Results of Negative Binomial Regression Using Dummy Variables of Local Market Structure (coefficient estimates)

Variable	Coefficient
StoresAbove0	-0.030* (0.013)
StoresAboveMedian	-0.051** (0.012)
StoresAbove30	-0.059** (0.014)

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

Once again, the coefficients on these dummy variables which capture variations in the local market structure are all negative and highly significant. After controlling for demographic and socioeconomic variables, urban area versus non-urban area effect, and historical purchasing measures, we find that the demand at the direct retailer is 3.0% lower for customers who have access to at least one store within 5 miles than for customers who do not have access to any stores within 5 miles. The coefficients on dummy variables *StoresAboveMedian* and *StoresAbove30* can be interpreted in similar ways.

We have also checked the robustness of our findings by estimating Probit and Logit models with these three dummy variables (instead of a negative binomial regression model), by using the number of stores within 10 miles, 15 miles, 25 miles, and 50 miles (instead of within 5 miles). Consistently we have found

a qualitatively similar result that the local market structure can have a statistically significant impact on the demand at the direct retailer. For the sake of brevity, we will not report these results.¹²

4.5. Testing for Endogeneity

Next we will test for endogeneity (or non-orthogonality), although previous related research has not expressed any concerns regarding endogeneity. This endogeneity (or non-orthogonality) concern arises when there may exist unobservable factors that affect both the number of local stores and the customer's demand. In such a case, the variable of the number of local stores would be correlated with (a.k.a. non-orthogonal to) the error term in our statistical model. To address this concern, we test for endogeneity in our model by using the total population in a zip code and the percentage of non-white population in a zip code as the instrument variables for the number of local stores. The total population and race of an area could be correlated with the local market structure (Goolsbee 2001; Sinai and Waldfoegel 2004). The total population and percentage of non-white population are assumed to be uncorrelated with the unobservable factors in our model for customer demand.

Following the two-step procedure for testing endogeneity that is outlined in Wooldridge (2001), we first regress *NumStores* onto other control variables and the two instrumental variables—the total population in a zip code in natural log and percentage of non-white population. Subsequently, we include the estimated residuals from this regression as an explanatory variable in our original Probit model. A statistically significant coefficient on the residuals would signal that *NumStores* is endogeneous, and a statistically insignificant coefficient on the residuals would signal that *NumStores* is not endogeneous. The estimates from such a two-step procedure are presented in column (1) of Table 6. Column (2) includes the estimates from the original model to aid comparison. The coefficient on the *Residuals* is not significantly different from zero. Thus, we fail to reject the null hypothesis that *NumStores* is exogenous.

¹² One could also 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 robust to using such a dummy variable as a control variable.

This result indicates that we can continue to use our original statistical models. Although this is not the purpose of this endogeneity-testing procedure, we note that the two-step procedure produces estimates that are very similar to the results using the original Probit model.

Table 6: Test for Endogeneity

	Endogeneity Test (1)	Original Probit (2)
NumStores	-0.018** (0.005)	-0.012** (0.003)
Recency	-0.039** (0.007)	-0.038** (0.007)
Frequency	0.232** (0.005)	0.232** (0.005)
Monetary Value	-0.028** (0.005)	-0.028** (0.005)
Median Income	0.084** (0.016)	0.083** (0.016)
Population Density	0.007 (0.008)	0.000 (0.006)
Median Age of Female	0.003** (0.001)	0.004** (0.001)
Percentage with Bachelor's Degree	0.035 (0.041)	0.013 (0.037)
Percentage Female Population	0.036 (0.200)	-0.041 (0.191)
Residuals	0.008 (0.006)	
Intercept	-1.921** (0.204)	-1.894** (0.203)
Log Likelihood	-82809.0	-82809.8
Sample Size	163,891	163,891

Standard errors are listed in parentheses.

** Significantly different from zero, $p < 0.01$

* Significantly different from zero, $p < 0.05$

4.6. Effect of Local Market Structure Varies across Different Products

In theory, the competition between direct retailing and brick-and-mortar retailing should also vary for different products. In the clothing industry, the runway shows and various fashion magazines play a key role in setting the current trend for popular designs (Aginis 1999). Subsequently, each clothing retailer would offer its own version of the products that are consistent with the fashion trend (Rantisi 2002).

According to our discussion with the senior management of the focal direct retailer, they also follow this industry-wide strategy when they design their products. This direct retailer offers products that follow the current style and color trends, and these products tend to be popular and have large sales. But this retailer also offers many products that do not follow the current trend and products with styles and colors that are more unique to the retailer.

Presumably, physical stores, with their more limited shelf-space and fewer SKUs, will focus on popular products. As a result of popular products being widely available in local stores, the competition between a direct retailer and local stores would be intense for popular products. On the other hand, a direct retailer is likely to face little or no competition if it sells niche products that are unlikely to be available in local stores. We will empirically analyze these theoretical predictions.

In our sample, there are 1,866 unique products that have positive sales.¹³ We rank the sales of all 1,866 products to identify the top bestselling products. Brynjolfsson et al. (2006) has shown that the widely-accepted Pareto Principle (also known as the 80/20 rule) can also be used to analyze the distribution of product sales. Thus, we define the top bestselling products that generate 80% of total sales as “popular products”.¹⁴ The rest of the products are defined as “niche products”. Subsequently, we estimate the effect of the number of local stores on the demand for popular and niche products separately. The results

¹³ Here we do not consider different colors as different products. If we consider different colors of the same item as different products, then we have 4,588 unique products. Our results are robust to considering each item-color combination as a unique product.

¹⁴ Our findings in this paper are robust to defining the top bestselling products that generate 50% of total sales as popular products.

are shown in Table 7. For popular products, the coefficient associated with *NumStores* is negative and significantly different than zero. In contrast, for niche products, the coefficient associated with *NumStores* is not statistically different from zero. Thus, the impact of the local market structure on customer demand through direct channels is almost entirely via popular products. Meanwhile, niche products stocked by the direct retailer are virtually immune from cross-channel competition.

Table 7: Negative Binomial Regression of Demand onto Local Market Structure

	Popular	Niche
NumStores	-0.026** (0.004)	0.006 (0.006)
Recency	-0.041** (0.011)	-0.311** (0.015)
Frequency	0.400** (0.008)	0.704** (0.012)
Monetary Value	-0.047** (0.009)	-0.248** (0.014)
Median Income	0.175** (0.025)	0.072 (0.037)
Population Density	-0.005 (0.011)	0.034* (0.013)
Median Age of Female	0.007** (0.001)	-0.001 (0.002)
Percentage with Bachelor's Degree	0.027 (0.060)	-0.032 (0.088)
Percentage Female Population	-0.132 (0.301)	-0.473 (0.437)
Intercept	-2.680** (0.326)	-0.893 (0.474)
Log Likelihood	-151,176.0	-66,771.0
Sample Size	163,891	163,891

Standard errors are listed in parentheses.

** Significantly different from zero, $p < 0.01$

* Significantly different from zero, $p < 0.05$

4.7. 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 Internet markets can significantly lower consumer search costs and a reduction in consumer search costs on the Internet can impact prices and price dispersion (e.g., Brynjolfsson and Smith 2000, Brown and Goolsbee 2002, Hann et al. 2003, Clay et al. 2003). Lower search costs on the Internet channel may also change the concentration of product sales. Brynjolfsson et al. (2006) 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. If the competition between direct retailing and brick-and-mortar retailing varies for different products, and if the Internet channel sells proportionately more niche products than the catalog channel, then it is possible that the level of competition between the Internet channel and local stores is different from the level of competition between the catalog channel and local stores.

So far we have only analyzed the effect of local stores on the overall customer demand at the direct retailer. In this section, we will study how the impact of local market structure varies across different direct retailing channels. We will conduct our analyses in three steps. First, we will analyze whether the Internet channel sells proportionately more niche products than the catalog channel. Second, we will study whether the effect of the local market structure varies across for different products at the Internet channel and whether this effect varies for different products at the catalog channel. Finally, we will use an equation to formally illustrate the intuition that the effect of the local market structure can vary across different channels, and we provide empirical results that are consistent with this intuition.

First, we compare the concentration of product sales through the Internet channel with that through the catalog channel. Following Brynjolfsson et al. (2003, 2006), we estimate a log-linear relationship between sales and sales rank and 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 the difference between these two slope parameters is highly significant, indicating that the sales distribution for the Internet channel is

significantly less concentrated than that for the catalog channel. Consistent with this result, 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, using the same definition of “popular products” and “niche products” as in Section 4.6. We interpret this as evidence that the Internet channel sells proportionately more niche products than the catalog channel.

Next we will repeat the same analysis as in Section 4.6, but we will analyze the Internet channel and the catalog channel separately. Once again, we use a negative binomial regression to estimate the impact of local market structure on the demand for different products. Table 8 reports the estimates for the catalog channel and for the Internet channel separately.

When considering the demand for popular products, the coefficient for the local market structure is negative and highly significant for both channels. This indicates that the demand for popular products declines as the number of local stores increases. This is consistent with the observation that popular products are likely to be stocked by local stores, and therefore the number of local stores can negatively impact the demand for such items. Interestingly, when considering the impact on demand for niche products, the coefficient on *NumStores* is statistically insignificant for both the catalog channel and the Internet channel. This suggests that the number of local stores does not negatively impact customers’ demand, through the Internet and catalog channels, for niche products which are unlikely to be widely available at local stores.

Finally, we will 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. The coefficient on the local market structure variable in the negative binomial regression model in equation (2) can be expressed as the marginal effect of *NumStores* on the total demand in natural log. More formally, we have

$$\beta = \frac{\partial \ln E(y_i | X_i)}{\partial x_i} \quad (3)$$

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 coefficient on the local market structure variable.

Total demand can be written as the sum of the demand for popular products (y_{iP}) and the demand for niche products (y_{iN}). That is,

$$y_i = y_{iP} + y_{iN} \quad (4)$$

Table 8: Negative Binomial Regression of Demand onto Local Market Structure for Different Channels

	Catalog Channel		Internet Channel	
	Popular	Niche	Popular	Niche
NumStores	-0.029** (0.006)	0.011 (0.009)	-0.020** (0.006)	0.003 (0.009)
Recency	0.085** (0.014)	-0.157** (0.022)	-0.189** (0.015)	-0.434** (0.019)
Frequency	0.274** (0.010)	0.554** (0.016)	0.509** (0.012)	0.797** (0.016)
Monetary Value	0.016 (0.011)	-0.143** (0.019)	-0.108** (0.013)	-0.321** (0.019)
Median Income	0.100** (0.032)	-0.066 (0.051)	0.270** (0.036)	0.246** (0.051)
Population Density	0.019 (0.013)	0.053** (0.018)	-0.034* (0.017)	0.014 (0.020)
Median Age of Female	0.014** (0.002)	0.005 (0.003)	-0.005* (0.002)	-0.007** (0.003)
Percentage with Bachelor's Degree	-0.241** (0.078)	-0.284* (0.126)	0.346** (0.085)	0.123 (0.120)
Percentage Female Population	0.466 (0.394)	0.312 (0.624)	-0.829 (0.433)	-1.272* (0.587)
Intercept	-3.664** (0.426)	-1.784** (0.669)	-2.928** (0.470)	-1.920** (0.647)
Log Likelihood	-100161.8	-37643.0	-81643.7	-40804.5
Sample Size	163,891	163,891	163,891	163,891

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

Substituting (4) into (3) gives us

$$\begin{aligned}
\beta &= \frac{\partial E(y_i | X_i)}{\partial x_i} \frac{1}{E(y_i | X_i)} \\
&= \frac{\partial (E(y_{iP} | X_i) + E(y_{iN} | X_i))}{\partial x_i} \frac{1}{E(y_i | X_i)} \\
&= \beta_P \frac{E(y_{iP} | X_i)}{E(y_i | X_i)} + \beta_N \frac{E(y_{iN} | X_i)}{E(y_i | X_i)}
\end{aligned} \tag{5}$$

where β is the estimate of the effect of local market structure on the total demand, β_P and β_N are the estimates of the effect of local market structure on the demand for popular products and for niche products, respectively.

Results in Table 8 show that, for both the Internet and catalog channels, the effect of local market structure on the demand for popular products is negative and significant, while the effect of local market structure 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 (5), we can conjecture that the effect of local stores on the total demand should be smaller in size for the Internet channel than for the catalog channel.

To test this hypothesis we analyze how the local market structure affects the total demand through these two channels separately. In particular, we replace the total demand in our previous negative binomial regression model first with the catalog demand and then with the Internet demand, and we estimate these two models separately. These results are reported in Table 9.

The results show that the coefficient on variable *NumStores* is negative and highly significant, when the catalog demand is used as the dependent variable. When the Internet demand is used as the dependent variable, the coefficient on variable *NumStores* is still negative but barely significant at the 5% significance level. This suggests that the catalog channel strongly competes with traditional 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 -0.021. We interpret this as evidence that the level of competition between the Internet channel and local stores is lower than the level of competition between the catalog channel and local stores, although the evidence is not particularly strong. To aid interpretation, we calculate how an increase of the number of local stores from 0 store (the 25th percentile of the sample distribution) to 7 stores (the median of the sample distribution) can impact the total demand. Ceteris paribus, such an increase in the number of local stores would reduce the total demand through the catalog channel by 4.4%, while it only reduces the total demand through the Internet channel by 2.5%.

Table 9: Negative Binomial Regression of Demand onto Local Market Structure for Each Channel

	Catalog	Internet
NumStores	-0.021** (0.005)	-0.012* (0.006)
Recency	0.062** (0.014)	-0.214** (0.014)
Frequency	0.280** (0.010)	0.510** (0.011)
Monetary Value	0.016 (0.010)	-0.115** (0.012)
Median Income	0.069* (0.031)	0.254** (0.034)
Population Density	0.026* (0.012)	-0.037* (0.016)
Median Age of Female	0.012** (0.002)	-0.006** (0.002)
Percentage with Bachelor's Degree	-0.242** (0.075)	0.298** (0.081)
Percentage Female Population	0.366 (0.374)	-0.714 (0.413)
Intercept	-2.959** (0.404)	-2.332** (0.446)
Log Likelihood	-109637.4	-92271.2
Sample Size	163,891	163,891

Standard errors are listed in parentheses.

** Significantly different from zero, $p < 0.01$

* Significantly different from zero, $p < 0.05$

5. Discussion and Implications

Our study shows strong evidence that geography has significant implications for direct retailing. In fact, an increase in the number of local stores can substantially reduce consumers' overall demand at a direct retailer. Prior theoretical research has suggested that the direct retailer strongly competes with local stores. Previous empirical research has found consumer demand through the Internet will be higher in local markets where local prices and sales tax rates are high. However, the impact of the local market structure, i.e., the number of local stores, on consumer demand through the direct channels that include the Internet and catalog channels has not been explored by previous literature. This research fills the void by empirically validating the assertion that as the number of local stores increases for a customer, the demand of the customer at a direct retailer decreases. The competition between local stores and direct channels we have identified in this paper becomes increasingly important, as the share of retailing revenue brought in by direct channels such as Internet and catalog channels continues to climb.

Moreover, we analyze the effect of the local market structure on the demand through each channel separately. Our findings advance the existing literature that has primarily emphasized how electronic market can improve consumer welfare through lower prices (Brynjolfsson and Smith 2000), improved product variety (Brynjolfsson et al. 2006; Brynjolfsson et al. 2003) and lower search costs (Bakos 1997). We find strong evidence that by greatly lowering search costs, the Internet is not only flattening sales distribution but also mitigating the competition with local stores. 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 catalog demand, whereas the impact of the local market structure on the Internet demand is smaller, a difference that can be attributed to lower search costs of the Internet. Strikingly, we find that higher proportion of demand for niche products softens the impact of the local market structure on the Internet channel. This finding emphasizes the key role of the wider selection available on the Internet in the substitution between online and offline channels.

Although the relationship between the Internet shopping and geography from the perspective of costs associated with distance (Forman et al. 2005; Sinai and Waldfogel 2004), convenience (Chiou 2005; Forman et al. 2006), and lower prices (Forman et al. 2006; Goolsbee 2000) has been recognized in the literature, the relationship between geography and product selection has garnered very little attention in ecommerce research. Given the magnitude of cross-channel substitution that we find, and the sharp differences across different products, this should be an area of continuing research.

Our research has important implications for both offline and direct retailers. We provide insights on how the local stores compete with the catalog channel as well as the Internet channel, a topic which has been under-explored. A direct retailer is more likely to have higher demand in areas that are under-served by local brick-and-mortar stores than in areas that have a high concentration of traditional stores. This effect exists even after controlling for demographic and socioeconomic variables, urban area versus non-urban area effect, and historical purchasing measures which are used to measure customer heterogeneity.

Furthermore, we find that popular products sold by a direct retailer strongly compete with local stores, while the competition is softened for niche products sold by a direct retailer. A direct retailer may want to develop a strategy of developing and marketing niche products, in order to win the battle with local stores. This strategy of selling niche products can be quite important and effective in helping a direct retailer win its battle with local stores, in particular when such a direct retailer sells to consumers who live in areas that have a high concentration of local stores. Overall, it is worthwhile for a direct retailer to vary its promotional strategies and product offerings based on the geographic location of the customer. For example, when customers move from areas under-served by local stores to areas well-served by local stores or vice-versa, a direct retailer needs to adjust its promotional strategies.

Information technologies in general and Internet markets in particular have lowered consumer search cost. As a result, the Internet channel sells proportionately more niche products than the catalog channel. This may soften the competition between a direct retailer and local stores. As the share of revenue brought in by a direct retailer's Internet channel grows, the competition between a direct retailer and local stores may

become less intense. This is good news for both direct retailers and local stores. Information technologies and Internet markets allow different types of retailers to pursue a differentiation strategy, which in the end may lead to less competition and potentially higher profitability for these retailers.

6. Conclusion

In this study, we develop insights by analyzing a unique combination of datasets on consumer purchases and their local market structures. Our focus is on the competition between local stores and direct retailing channels. In particular, after controlling for consumer differences, we examine whether consumers with few local stores shop more from a direct retailer than do consumers with multiple local retailing options. In addition, we identify the role of demand for popular products and demand for niche products in shaping the competition between local stores and the Internet and catalog channels—Internet sales of niche products, which are often unavailable in physical stores, are largely immune from competition by traditional retailers.

We present significant evidence that as the number of local stores increases a consumer's demand at the direct retailer decreases, a finding consistent with the predictions of economic theory and prior theoretical research. We also show that the competition between local stores and the direct retailer is primarily geared toward popular products. The competition between local stores and the direct retailer is muted when the direct retailer sells niche products, products that are typically not stocked by local stores.

More importantly, we demonstrate that niche products account for a larger proportion of customer purchases through the Internet channel than through the catalog channel, an outcome ensuing from the low search costs on the Internet and consistent with prior research. This relatively higher demand for niche products in turn mitigates the competition of the Internet with local stores. This provides substantive indication of product selection effect in online and offline competition.

In general, our findings suggest that businesses must consider geography and local market structure variables in strategically targeting customers. For instance, online retailers should emphasize greater

product selections to consumers with many traditional stores nearby. Thus, marketing communications must accommodate customers' geographic heterogeneity.

References

- Agins, T. "The Real Action In Fashion Is Off the Runway," in: *Wall Street Journal*, New York, N.Y., Jul 30, 1999, p. B.1.
- Anderson, C. "The Long Tail " in: *Wired Magazine*, 2004.
- Anderson, E., and Simester, D. "Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies," *Marketing Science* (23:1) 2004, pp 4-20.
- Bakos, J.Y. "Reducing buyer search costs: Implications for electronic marketplaces," *Management Science* (43:12) 1997, pp 1676-1692.
- Balasubramanian, S. "Mail Versus Mall: A Strategic Analysis of Competition Between Direct Marketers and Conventional Retailers," *Marketing Science* (17:3) 1998, pp 181-195.
- Bitran, G.R., and Mondschein, S.V. "Mailing Decisions in the Catalog Sales Industry," *Management Science* (42:9) 1996, pp 1364-1381.
- Bradlow, E., Bronnenberg, B., Russell, G., Arora, N., Bell, D., Duvvuri, S., Hofstede, F., Sismeiro, C., Thomadsen, R., and Yang, S. "Spatial Models in Marketing," *Marketing Letters* (16:3/4) 2005, pp 267-278.
- Bresnahan, T.F., and Reiss, P.C. "Entry and competition in concentrated markets," *Journal of Political Economy* (99:5) 1991, p 977.
- Brown, J. R., and Goolsbee, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy* (110:3) 2002, pp 481–507.
- Brynjolfsson, E., Hu, Y.J., and Simester, D. (2006). "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales." Working Paper, MIT Sloan School of Management.

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

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

Cameron, A.C., and Trivedi, P.K. *Regression Analysis of Count Data* Cambridge University Press, 1998.

Campbell, J.R., and Hopenhayn, H.A. "Market size matters," *Journal of Industrial Economics* (53:1), Mar 2005, pp 1-25.

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

Clay, Karen, Ramayya Krishnan, Eric Wolff, and Danny Fernnades. 2003. Retail Strategies on the Web: Price and Non-Price Competition in the Online Book Industry. *Journal of Industrial Economics* 49(4) 521-540.

Dubbs, D. "Mind Games," in: *Multichannel Merchant Magazine*, Stamford, CT, Jan 1, 2002.

Forman, C., Ghose, A., and Goldfarb, A. (2006). "Geography and Electronic Commerce: Measuring Convenience, Selection, and Price." Working Paper, New York University.

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

Glaeser, E.L., Kolko, J., and Saiz, A. "Consumer city," *Journal of Economic Geography* (1:1) 2001, p 27.

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

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

Greene, W.H. *Econometric Analysis*, (5th ed.) Prentice Hall, 2002.

Hann, Il-horn, Eric Clemons, and Lorin Hitt. 2003. Price Dispersion and Differentiation in Online Travel: An Empirical Investigation. *Management Science* 48(4) 534-549.

Hausman, J., Hall, B.H., and Griliches, Z. "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica* (52:4) 1984, pp 909-938.

Jank, W., and Kannan, P.K. "Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice," *Marketing Science* (24:4), Fall 2005, pp 623-634.

Prince, J. "The Beginning of Online/Retail Competition and Its Origins: An Application to Personal Computers," *International Journal of Industrial Organization* (25:1) 2007, pp 139-156.

Rantisi, N.M. "The competitive foundations of localized learning and innovation: The case of women's garment production in New York City," *Economic Geography* (78:4) 2002, p 441.

Rubin, M. "Holiday E-Commerce Spending Up 26 Percent With Surge Continuing the Week Before Christmas," in: *PR Newswire*, Dec 28, 2006.

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

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data* The MIT Press, 2001, p. 740.