AN ANALYSIS OF PRICING STRATEGY AND PRICE DISPERSION ON THE INTERNET

Randy A. Nelson
Colby College

Richard Cohen
Buckingham Properties

and

Frederik Roy Rasmussen
Harvard Business School

INTRODUCTION

The development of e-commerce, or the buying and selling of goods on the internet, has greatly reduced the costs to consumers of identifying the sellers of various goods and the prices that they charge. This is particularly true for consumers who use "price robots" or "shopbots" such as pricescan.com, mysimon.com, bestbookbuys.com, cnet.com, etc[1]. Shopbots allow consumers to define functionally equivalent products, such as a laptop computer with a given microprocessor, hard disk space, memory, etc., or a specified make and model of a particular product; the shopbot then identifies vendors who sell the product and provides a price quote for each vendor. By reducing search costs, many analysts expected shopbots (and the internet in general) to make electronic markets more competitive, leading to lower prices and less price dispersion[2]. On the other hand, theoretical work by Stahl [1989] and Chen and Hitt [2003] indicate that a reduction in search costs may lead to an increase in price dispersion, an hypothesis that is supported in part by the empirical work of Brown and Goolsbee [2002]. Finally, Baye and Morgan [2004] demonstrate that price dispersion may be an equilibrium outcome if firms operate subject to bounded rationality resulting from satisficing behavior, or if firms are reluctant to alter pricing strategies because they view the payoffs as small relative to the costs.

Early studies of price dispersion on the internet, which focused primarily on books and compact disks (CDs), have produced mixed results. Bailey [1998] computed the standard deviation of prices for books, CDs and computer software for both internet and traditional retailers, reporting that price dispersion was greater on the internet for books and CDs, but lower for computer software. Brynjolfsson and Smith [2000] report the results of a similar experiment in which they found that the standard deviation of prices for internet sellers was greater than that of traditional retailers 84 percent of the time for books, but only 33 percent of the time for CDs[3]. A similar

Randy A. Nelson: Department of Economics, Colby College, Waterville, Maine 04901. E-mail: ranelson@colby.edu.

study by Friberg, Ganslandt, and Sandstrom [2000] found that the range of prices for books and CDs in Sweden was generally greater for internet sellers than for their bricks and mortar competitors. In two recent studies of price dispersion for books sold online, Clay, Krishnan, and Wolff [2001] found that the coefficient of variation of prices was greater for bestsellers than for randomly chosen books, and lower for hardcover books relative to paperbacks, while Clay, Krishnan, Wolff, and Fernandes [2002] report that the average prices for books were similar in online and bricks and mortar stores, and that substantial price dispersion existed online.

More recent studies have attempted to explain why price dispersion continues to exist in on-line shopping, focusing on seller characteristics (Pan, Ratchford, and Shankar [2002, 2003], Ratchford, Pan, and Shankar [2003], Clay, Krishnan, Wolff, and Fernandez [2002], and Baye, Morgan, and Sholten [2004]), the number of sellers (Baye, Morgan, and Sholten [2003], Ratchford, Pan, and Shankar [2003], and Clay, Krishnan, and Wolff [2001]), and time of entry (Pan, Ratchford, and Shankar [2003]).

In this study we expand on the previous literature dealing with price dispersion on the internet in several dimensions. Previous studies have typically focused on a small number of products such as books, CDs, computer hardware or software, etc.; in this study we begin by presenting evidence on the degree of price dispersion for 542 homogeneous products in 13 different broadly defined product categories. Second, numerous theories have been put forth to explain why price dispersion may exist in competitive markets; we test several of these theories by exploring the relationship between price dispersion and the number of sellers, average price, and frequency of purchase. Third, we test to determine whether the level of price dispersion changes over time, and examine the impact of entry on changes in price dispersion over time. Fourth, we explore the impact of shipping and handling charges and seller heterogeneity on price dispersion. Finally, based on Stigler's [1961] argument that search costs will be reduced if prices are correlated over time, we investigate whether the low price sellers for a particular product or at a point in time remain among the low price firms for similar products or future time periods.

PRICE DISPERSION

Data

To obtain the necessary data five different shopbots were used to collect price quotes for 542 products in thirteen different product categories. The pricebots bestbookbuys.com and nextag.com were used primarily to obtain price information on books, pricescan.com and cnet.com for consumer electronics, and mysimon.com for music. Although it is possible to obtain price data on a particular product from more than one pricebot, in each case only a single pricebot was used to obtain price information for a given good. In each case the price quotes are for a unique, homogeneous product, representing either a specific make and model of a good such as a computer or digital camera, or a specific title for books and CDs. A total of 7,519 price quotes were obtained during January of 2000; the minimum number of price quotes for any good was two, the maximum 59, the average 13.87. For each product the average

price, standard deviation, and coefficient of variation (the standard deviation divided by the average price) was computed. The means of each of the variables, weighted by the number of price quotes, is reported for each product category in Table 1.

TABLE 1
Descriptive Statistics

Product Category	Number of Products	Weighted Average Price	Weighted Average Standard Deviation	Weighted Average Coefficient of Variation	Coefficient of Variation: Range	Average Number of Sellers
Watches	53	$2023.56	$119.62	12.25%	.11%-3.89%	3.26
Copiers	22	$628.25	$108.12	17.67%	.01-41.15%	13.32
Games	25	$41.40	$4.00	10.31%	.03-33.85%	9.04
Hardcover Books	62	$22.70	$4.17	18.64%	1.38-29.65%	10.82
Paperback Books	55	$12.42	$1.94	15.92%	1.47-25.11%	9.69
CDs	56	$13.57	$1.64	12.14%	4.14-26.66%	11.14
Portable Audio Equipment	32	$180.32	$21.55	13.42%	4.46-37.46%	10.59
Home Audio Equipment	55	$312.62	$43.89	13.14%	1.62-45.27%	8.98
Personal Digital Assistants	9	$377.96	$50.26	12.75%	5.77-24.63%	30.89
Computers/Monitors	48	$2605.76	$162.14	6.72%	.01-53.20%	13.77
Cameras	42	$699.13	$56.55	8.59%	.06-65.20%	28.26
Computer Printers/Scanners	41	$367.04	$33.58	9.51%	3.48-57.90%	26.56
Computer Accessories	42	$165.24	$16.22	10.63%	4.43-34.69%	22.64
Overall Weighted Average*	41.69	$532.46	$43.54	11.69%		13.87

*The averages for the number of products and number of sellers are simple, and not weighted, averages.

The weighted average prices range from a low of $12.42 for paperback books to a high of $2,605.76 for computer/monitors, with an overall mean of $532.46. The overall weighted average standard deviation is $43.54, with a minimum of $1.64 for CDs and a maximum of $162.14 for computers/monitors. The weighted average coefficients of variation range from a low of 6.72 percent for computers and computer monitors to a high of 18.64 percent for hardcover books; the average is 11.69 percent. The coefficients of variation for hardcover books (18.64 percent), paperback books (15.92 percent), and CDs (12.14 percent) are similar to those reported by Bailey [1998] for books (13.19 percent) and CDs (17.61 percent) available online, and to those for hardcover (15-18 percent) and paperback (10-14 percent) books reported by Clay, Krishnan, and Wolff [2001].

Price Dispersion and Consumer Search

Starting with Stigler [1961], a number of theories have been put forth to determine the factors that explain consumer search behavior, and ultimately, the level of price dispersion. We use the comparative-static results from these models to develop a simple regression model to explore the determinants of price dispersion. In the search models developed by Stigler [1961], Carlson and McAfee [1983], Dana [1994], and Janssen and Moraga [2000], the level of price dispersion is positively related to the number of suppliers. To test for the effects of competition on the level of price dispersion we include the variable *SELLERS* in our model, which is defined to be the number of price quotes obtained for each product. Carlson and McAfee [1983] dem-

onstrate that higher costs lead to both higher prices and an increase in the variance of prices. To test this hypothesis the variable *PRICE*, defined to be the average price of the product, is included in the regression.

Finally, Stigler [1961] has argued that consumers have a greater incentive to search for price information for products that are subject to repetitive purchases relative to products that are purchased on an infrequent basis. As a result, one would expect to find an inverse relationship between the frequency of purchase and the level of price dispersion. Unfortunately, we did not have data for the frequency of purchase for any of the products included in our sample. As an alternative, a dummy variable *FREQUENT* is employed, which takes on a value of 1 for the four products in our sample, computer games, hardcover books, paperback books, and CDs that would normally be purchased several times a year, and 0 otherwise. Although consumers would not purchase the same book or CD title more than once, they would have an incentive to identify the lowest price seller for a given title if they intended to purchase additional titles in the future and if they believed that the low price seller for one title would also be the low price seller for additional titles[4].

The inclusion of the variable *FREQUENT* precludes the use of a fixed-effects panel model with product category dummy variables, as the result would be a singular data matrix. We thus estimated two different models, one including *FREQUENT* but no product dummy variables, and one including product category dummy variables but not *FREQUENT*. The two models may thus be written as:

$$(1) \quad \ln STDEV = \alpha + \beta_1 \ln PRICE + \beta_2 \ln SELLERS + \beta_3 FREQUENT + \varepsilon$$

$$(2) \quad \ln STDEV = \alpha + \sum_{i=1}^{12} \delta_i D_i + \beta_1 \ln PRICE + \beta_2 \ln SELLERS + \varepsilon$$

where *STDEV* is the standard deviation of prices and ln is the natural log operator[5].

Given the cross-sectional nature of the data, the models were estimated using the White [1980] estimator to obtain heteroskedasticity-robust standard errors. Parameter estimates for both models are presented in the first two columns of Table 2. The estimated parameters of ln*PRICE* and ln*SELLERS* are both of the predicted sign, statistically significant at the 1 percent confidence level, and are of similar magnitude in both models. The average (across the two models) elasticity of *STDEV* with respect to *PRICE* is .6973, a result that is consistent with the positive relationship between the standard deviation of prices and price reported by Pratt et al. [1979], Dahlby and West [1986], Sorensen [2000], and Clay, Krishnan, and Wolff [2001]. It should be noted, however, that this relationship is inelastic, implying that the coefficient of variation is negatively related to price; dispersion thus decreases as a percentage of price as the price increases[6]. The elasticity of *STDEV* with respect to *SELLERS* is .2777, implying that an increase in the number of sellers across markets results in greater price dispersion; this finding is consistent with that reported by Clay, Krishnan, and Wolff [2001], who found that the standard deviation of prices increased as the number of booksellers increased. Pan, Ratchford, and Shankar [2003], however, found that the standard deviation of prices is inversely related to the number of sell-

ers, while Baye, Morgan, and Scholten [2003, 2004] found that the gap between the lowest and average or gap between the lowest and second lowest prices charged in a given market, respectively, are inversely related to the number of sellers. Finally, the coefficient of variation for products that would normally be purchased several times a year is 52.04 percent lower than products that would typically be purchased on an infrequent basis. This result is consistent with the results presented by Sorensen [2000] for prescription drugs, and with hypothesis that consumers have a greater incentive to search for frequently purchased goods because the information obtained may be useful for future purchases.

TABLE 2
Regression Results

	Model I		Model 2		Model 3		Model 4	
Variable	**Coeff.**	**t-stat**	**Coeff.**	**t-stat**	**Coeff**	**t-stat**	**Coeff**	**t-stat**
Constant	-.8929	3.74	-2.1178	4.56	6.6374	13.32	10.1024	6.29
lnPRICE	.6710	14.38	.7237	12.39	-.6635	6.61		
lnSELLERS	.1845	2.70	.3709	3.30	-.0572	2.91		
FREQUENT	-.7347	4.94						
WEEK					.0015	2.15		
Watches			1.0162	3.90				
Copiers			.2548	0.56				
Games			-.4145	1.43				
Hardcover Books			.2376	1.42				
Paperback Books			-.0920	0.45				
Portable Audio Eq.			.8292	4.75				
CDs			-.2527	1.37				
Home Audio Eq.			.9426	5.33				
PDAs			.2069	1.03				
Computers/Monitors			.1001	0.30				
Cameras			.1441	0.93				
Printers/Scanners			.3208	1.82				
SHIPPING							-.5646	4.41
SELECTION							-.3768	1.47
REPRESENTATION							1.2565	6.69
INFORMATION							.7480	3.14
TRACKING							-1.0816	7.13
$\overline{R}^2$	.70		.72		.68		.83	

Price Dispersion Over Time

Controlling for the number of sellers and the average price of the product, we hypothesize that the level of price dispersion will decline over time for two reasons. First, as consumers begin to identify the low cost sellers and adjust their purchases accordingly, the high priced firms will have an incentive to exit the market or adjust their prices downward. Given the low menu costs for internet based sellers it should be relatively easy and inexpensive for firms to make the necessary price adjustments[7]. Second, in most models of competition or oligopoly the entry of new firms is thought to drive the market price towards the competitive level, thus reducing price dispersion. Given the relatively low entry barriers faced by e-tailers, we conjecture that the

entry of new firms would thus reduce the level of price dispersion over time. Baye, Morgan, and Scholten [2006], however, have argued that persistent price dispersion is a result that is consistent with a number of theoretical models of search or oligopoly, leading them to argue that price dispersion may be a long run equilibrium outcome in e-markets. To test these hypotheses we collected weekly price data for a sample of forty-eight products for a period of six months covering February through July of 2000. The sample covers a range of products typically sold on the internet, including seven audio systems, six DVD players or VCRs, nine computer accessories, eight books, five music CDs, four cameras, five watches, and four computer games. A total of 26,078 price points were collected, yielding a data set consisting of 1,198 observations, where each observation represents the standard deviation of the prices for a given product in a given week. The average number of sellers for a given product in a typical week was 21.77, with a minimum of 2 and a maximum of 68.

In an effort to examine how the level of price dispersion changes over time we estimate a model similar to models (1) and (2). To control for changes in the number of sellers and the average price of the product over time we include the variables ln*SELLERS* and ln*PRICE,* respectively. To allow for changes in the level of price dispersion over time we include the variable *WEEK*, representing the week in which the observation was recorded. The model may thus be written as:

$$\ln STDEV = \alpha + \sum_{i=1}^{45} \delta_i D_i + \beta_1 \ln PRICE + \beta_2 \ln SELLERS + \beta_3 WEEK + \varepsilon \quad (3)$$

The model was estimated using GLS, correcting for item-specific autocorrelation within panels, and cross-sectional heteroskedasticity across panels; parameter estimates are presented in the third column of Table 2.

The estimated coefficient for *WEEK* is positive and significant at the 1 percent level, implying that the level of price dispersion increases over time; holding the average price and number of sellers constant, the standard deviation increases by 3.84 percent moving from the first week of our sample to the last (week 25). Clay, Krishnan, Wolff, and Fernandes [2002] report that the standard deviation of a large sample of book prices declined slightly over the period August 1999 to January 2000, Ratchford, Pan, and Shankar [2003] found that price dispersion declined significantly between 2000 and 2001, while Baye, Morgan, and Scholten [2006] report no change in the gap between the lowest and next lowest price or the coefficient of variation of prices over an eight month period. The time series estimates of *lnPRICE* and *lnSELLERS* are both negative and significant at the 1 percent level, results that are the opposite of those obtained from the cross-sectional data. The cross-sectional results indicate that moving from markets with a low average price or a small number of sellers to markets with high average prices and a large number of sellers results in an increase in price dispersion; the time series results indicate that increasing the number of sellers or average price level within a given market over time decreases the level of price dispersion. This latter finding is consistent with the results presented in Baye, Morgan, and Scholten [2006], who found an inverse relationship between the number of firms and the coefficient of variation over time.

To further explore the effects of entry on price dispersion we compute the ratio of the average price charged by firms that entered after the first week relative to the average price charged by firms that were in the sample beginning in week one. Not surprisingly, the entering firms charge prices below those charged by the early movers; on average, entering firms charge prices that are 5.1 percent below those charged by firms in the sample in the first week. In addition, the standard deviation of prices charged by the entrants is 36.2 percent below that of the original firms. It thus appears that the prices charged by entering firms are concentrated in, and may extend the left-hand tail of the price distribution, thus increasing the overall level of price dispersion over time.

POSSIBLE EXPLANATIONS FOR PRICE DISPERSION

Many early analysts predicted that the widespread use of the internet in general, and shopbots in particular, would lead to increased competition among internet sellers, eventually forcing all firms to charge the competitive price. In spite of these optimistic predictions, past studies have found little difference between the level of price dispersion for e-tailers relative to traditional bricks and mortar stores, and that the level of price dispersion on the internet does not appear to consistently diminish over time. In this section we explore two possible explanations for this phenomenon.

Shipping Costs

The price data used to compute the results presented in Table 1 do not take into account shipping costs. Internet firms may attempt to attract customers with low product prices, and then compensate by imposing higher than average charges for shipping and handling. If shipping costs are negatively correlated with the price of the product, then delivered prices may converge across sellers, reducing the level of price dispersion[8]. To test this hypothesis data were collected for product prices and shipping charges for the three most popular goods sold on the internet, books, DVDs and CDs. In each case the first step was to obtain a list of the most popular items for each product class; a total of thirty titles were selected from the *New York Times*' best seller lists, fifteen fiction and fifteen non-fiction, and twenty titles each for DVDs and CDs from the best seller lists of several of the firms that sold these products. All of the firms surveyed were covered by Bizrate.com, a firm that ranks the quality of internet sellers. A total of 287 observations were obtained from the ten firms that sell books, 225 observations from thirteen firms that sell DVDs, and 234 observations from twelve firms selling CDs.

Sample means for product price, shipping and handling charges, and total price (product price plus shipping and handling), together with the mean values of the standard deviations and coefficients of variation for the product prices and total prices are presented in Table 3. The average shipping charges for books, DVDs, and CDs represent 20.22 percent, 13.80 percent, and 22.09 percent, respectively, of the total price. The range of shipping charges is $3.47 to $5.00 for books, $1.99 to $4.39 for DVDs, and $2.00 to $3.95 for CDs. The correlation coefficient between the product price and shipping charge is -.053 (p = .370) for books, -.117 (p = .079) for DVDs,

and -.124 (p = .057) for CDs. The negative correlation coefficients lend support to the hypothesis that firms compensate in part for low product prices by charging higher fees for shipping and handling.

TABLE 3
Average Prices and Shipping Charges

Variable	Books	DVDs	CDs
Product Price	$17.16	$22.25	$13.76
Shipping & Handling	$4.35	$3.07	$3.04
Total Price	$21.51	$25.33	$16.81
Standard Deviation Product Price	$2.21	$1.68	$1.17
Standard Deviation Total Price	$2.23	$1.59	$1.19

The inclusion of shipping charges produces a slight increase in the standard deviation for books and CDs, but neither difference is significant at the 10 percent level using a matched-pairs t-test. For DVDs, however, the standard deviation for the total price is 5.35 percent less than the standard deviation for the product price, a difference that is significant at the 1 percent level.

Seller Heterogeneity

Stigler [1961, 214] argues that "Price dispersion is a manifestation – and, indeed, it is the measure – of ignorance in the market. Dispersion is a biased measure of ignorance because there is never absolute homogeneity in the commodity if we include the terms of sale within the concept of the commodity. Thus, some automobile dealers might perform more service, or carry a larger range of varieties in stock, and a portion of the observed dispersion is presumably attributable to such differences." If web sites differ in terms of ease of use, product selection, or consumer confidence, and if consumers are willing to pay more for these attributes, then prices may differ across web sites for homogeneous products even if search costs approach zero.

Several studies, including those by Clay, Krishnan, Wolff, and Fernandes [2002], Pan, Ratchford, and Shankar [2002, 2003], Baye, Morgan, and Scholten [2003], and Ratchford, Pan, and Shankar [2003] have explored the relationship between various aspects of seller quality and prices. In general, these studies have (i) failed to agree on which attributes are most important in explaining price differences, (ii) demonstrated that a given attribute may impact different products in different ways, and (iii) that differences in e-tailer attributes explain little, if any, of the differences in the prices charged by different firms. To explore the impact of seller heterogeneity on price dispersion, rankings of internet sellers were obtained from Bizrate.com. This site provides rankings of e-commerce firms in terms of the quality of their web site (ease of ordering, product information, and web site navigation and looks), product selection (breadth/depth of products offered), product representation (product description vs. what you received), and product delivery (expected vs. actual delivery date, customer support, and product tracking). The price quotes for books, CDs, and DVDs used to examine the impact of shipping costs on price together with the BizRate rankings, provide the data necessary to estimate hedonic price regressions to determine the

extent to which seller attributes influence prices. Using the nominal and quality adjusted prices, it is then possible to determine the extent to which seller heterogeneity affects the level of price dispersion.

The three rankings dealing with both web site quality and product delivery are all highly correlated; the average of the pair-wise correlations coefficients is in excess of .9 in both cases. In an effort to minimize the impacts of multicollinearity we include only one ranking from each group in the model[9]. To control for the negative relationship between product price and shipping charges we include the cost of shipping and handling in the estimated equation. The fixed-effects panel model, with product-specific dummy variables, may thus be written as

$$P_{ij} = \alpha + \sum_{i=1}^{69} \gamma_i D_i + \beta_1 SHIPPING + \beta_2 SELECTION + \beta_3 REPRESENTATION + \beta_4 INFORMATION + \beta_5 TRACKING + \varepsilon_{ij} \quad (4)$$

where P_{ij} is the price of the ith item sold by the jth firm, *SHIPPING* is the charge for shipping and handling, and *SELECTION, REPRESENTATION, INFORMATION,* and *TRACKING* are the firm's ranking for production selection, product representation, product information, and product tracking, respectively.

Equation (4) was estimated using White's [1980] procedure to obtain robust standard errors; estimated parameters, obtained by pooling the data for all three products, are presented in the fourth column of Table 2. Previous attempts to explain the relationship between e-tailer prices and characteristics have generally reported low adjusted r-squares, implying that seller characteristics explain little if any of the variation in prices across sellers. Our results indicate greater success in this respect, in that we report an adjusted r-square of .83, indicating that differences in seller attributes explain a significant fraction of the variation across sellers for the three products we examine. Consistent with the results presented in the section dealing with shipping costs, the estimated coefficient for *SHIPPING* is negative and significant at the 1 percent confidence level. Controlling for other factors, firms reduce their product prices by an average of $.56 for each $1 increase in the cost of shipping and handling. The estimated coefficients for product representation and product information are both positive and significant at the 1 percent level, while the estimated coefficients for product selection and product tracking are both negative.

It is possible to correct for the effects of seller heterogeneity using the estimated parameters from equation (4). For example, firm j may be able to command a higher price because it has a higher ranking for attribute X_k than the average firm. Define P_{ij}^{adj} to be the adjusted price firm j would charge for product i if its rankings for the attributes discussed above were equal to the sample means. P_{ij}^{adj} may thus be computed as

$$P_{ij}^{adj} = P_{ij} - \sum \beta_k (X_{jk} - \bar{X}_k) \quad (5)$$

where $\bar{X}_k$ is the sample mean of attribute k. Firms with above average seller rankings (assuming $\beta_k > 0$) thus have their prices adjusted downward, and vice versa.

The standard deviation declines from $2.21 to $2.13 (3.60 percent) for books, from $1.68 to $1.63 (2.96 percent) for DVDs, and from $1.17 to $1.07 (8.89 percent) for CDs when calculated with the adjusted prices. Using a matched-pairs t-test, the null hypothesis that the means of the standard deviations for the nominal and adjusted prices are equal may be rejected at the 5 percent confidence level or better in every case; accounting for seller heterogeneity thus reduces the standard deviations by an average of 5.15 percent. This finding is consistent with the results presented in Brynjolfsson and Smith [2000], Clay, Krishnan, Wolff, and Fernandes [2002], Pan, Ratchford, and Shankar [2002, 2003], and Ratchford, Pan, and Shankar [2003], all of whom report that seller attributes contribute little to one's ability to explain differences in prices across firms.

PRICING PATTERNS ACROSS PRODUCTS AND OVER TIME

An important factor in determining consumer search costs in both e-commerce and traditional markets is the consistency of a firm's pricing strategy across products and over time. Search costs would be greatly reduced if consumers knew that a firm that had the lowest price for a given product or at a given point in time, was also the lowest price seller for all items in a given category or at all future points in time. Stigler [1961, 218] has argued that "As a rule, positive correlations (of prices) should exist with homogeneous products. … A seller who wishes to obtain the continued patronage of those buyers who value the gains of search more highly or have lower costs of search must see to it that he is quoting relatively low prices. In fact, goodwill may be defined as continued patronage by customers without continued search (that is, no more than occasional verification)."[10]

Varian [1980], however, has argued that firms have an incentive to vary their prices over time by holding periodic sales. This strategy allows sellers to price discriminate between informed customers, who know the entire distribution of offered prices and always select the low-priced firm, and uninformed consumers, who are ignorant of the distribution of prices and randomly select sellers. The sale-induced price fluctuations prevent consumers from learning by experience which firms are the low price sellers, thus creating "temporal" price dispersion.

This section examines pricing patterns across products and over time to determine whether firms with the lowest (highest) prices for a given product or at a given point in time maintain their low (high) price strategy for all products or time periods.

Pricing Patterns Across Products

The data collected for books, DVDs, and CDs discussed in the section dealing with shipping costs is used to examine pricing patterns across products. This data set contains price information for 30 hardcover books sold by 10 different firms, and 20 DVDs and CDs sold by 12 different firms each. For each of the book titles, for example, we rank the sellers from lowest price (1) to highest price (10); ties are assigned the average of the rankings that would have been assigned had no ties occurred. For each firm an average ranking is then computed across all products. If one firm is always

the lowest (highest) ranked firm for all titles, its average ranking will be 1 (10); if the price rankings are random across titles, the average ranking for all firms will be 5.5. Using Kendall's coefficient of concordance, W, it is possible to test the null hypothesis that the rankings are independent across products.[11] The average rankings and test statistics for all three product classes are reported in Table 4.

TABLE 4
Price Rankings Across Products

Seller	Books	DVDs	CDs
Firm 1	8.13	5.00	6.38
Firm 2	6.68	7.73	11.21
Firm 3	3.95	6.62	7.97
Firm 4	4.55	11.19	6.35
Firm 5	3.45	9.38	2.41
Firm 6	4.00	4.00	3.53
Firm 7	4.65	8.00	9.76
Firm 8	8.30	3.73	4.97
Firm 9	6.22	4.73	4.97
Firm 10	5.07	4.19	6.35
Firm 11		9.35	5.53
Firm 12		4.08	8.56
N	20	13	17
Kendall's W	.385	.524	.503
$\chi 2$ statistic	69.327	74.930	94.081
Significance Level	.000	.000	.000

N represents the number of products used to compute the average rankings; products not carried by all firms are dropped from the sample.

For all three product categories there are clear differences in the average rankings across sellers, indicating that some firms are consistently among the lowest price firms for all products, while others are consistently among the highest priced firms. Kendall's W is approximately distributed as χ^2 with degrees of freedom equal to the number of sellers minus one; in every case the null hypothesis that the rankings are independent across products may be rejected at the 1 percent confidence level. The results obtained using Kendall's W understate the strength of the pricing relationship across products if there are a large number of ties, or firms charging the same price. For example, if the lowest price for a given product is offered by five different firms, each firm's price ranking will be (5+4+3+2+1)/5 = 3, not 1. Firms may charge identical prices if they employ a form of markup pricing in which they offer the product at a fixed discount relative to the list price. Since the list price for books, CDs, and DVDs is readily available on many web sites, it is possible to compute the discount relative to the list price offered by each firm.

Rounding to the nearest whole percent, we find that 46.0 percent of the books in our sample were offered at a discount of 40 percent off the list price, 19.5 percent at a discount of 30 percent, and 13.2 percent at a discount of 20 percent; the remaining books were sold at discounts ranging from 50 percent to 0 percent. Five of the firms in the sample charged the same price, representing a discount of 40 percent from the list price, for 18 of the 30 titles; in every case this price ranked as the lowest or second lowest price offered for the book. Four of these firms charged the same price for an

additional 6 books, again at a discount of 40 percent from the list price. No firm employed a single 'discount rate' for all of the books included in our sample; the average number of discount rates was 4, with a minimum of 2 and a maximum of 8.

The pricing patterns for firms that sell CDs and DVDs are generally similar to those found for books. For the CDs in our sample, 29.1 percent were sold at a discount of 30 percent off the list price, 19.2 percent at a discount of 26 percent, and 6.0 percent and 6.8 percent at discount rates of 22 percent and 16 percent, respectively. The most common discount rates for DVDs was 30 percent, which accounted for 25.8 percent of our sample; discount rates of 25 percent and 20 percent account for an additional 9.8 percent and 17.8 percent of the sample, respectively.

Pricing Patterns Over Time

The data employed in the section dealing with price dispersion over time consisting of weekly prices for forty-six products collected over a six month period, were used to test the hypothesis that firms maintain their relative price rankings over time. A total of 25,778 price points were collected, covering seven audio systems, six DVD players or VCRs, eight computer accessories, six books, six music CDs, four cameras, five watches, and four computer games.

For each product we derived weekly price rankings, assigning values of 1 for the lowest price and 10 for the highest; ties were awarded the average of the rankings had no ties occurred. For each firm an average ranking across all time periods was then computed. Using Kendall's test of concordance it is possible to test the null hypothesis that the rankings are independent across firms. One drawback to the Kendall test is that it deletes the data for any week in which one or more firms are not represented[12]. In an effort to circumvent this shortcoming, only those firms with price quotes in at least 13 of the first 15 weeks were included in sample, a total of 628 firms[13]. The null hypothesis that the price rankings were independent over time was rejected at the 1 percent confidence level for every one of forty-six products, indicating the existence of "price persistence" over time for this subset of the data.

The results discussed above ignore the role that entry and exit play in determining relative price rankings over time. The number of firms may fluctuate as a result of a change in the list of firms covered by a given shopbot, the entry of new firms, the exit of existing firms, or the decision by existing firms to alter their product mix. To allow for the effects of entry and exit on the relative price rankings weekly price quartiles were constructed for the forty products in our sample with twelve or more sellers in every time period. For each firm the price quartile for the week in which the firm first and last appeared in the sample were recorded; this exercise was repeated for all firms that appeared in the sample a minimum of 5, 10, 15, 20 and 25 weeks. The results of this analysis are reported in Table 5.

The results presented in Table 5 indicate, for example, that 63.55 percent of the firms that started in the lowest price quartile and were in the sample for 5 or more weeks were also in the lowest quartile when they exited the sample. If firms were indifferent about maintaining their relative price rankings over time, one would expect to find a uniform distribution of prices across ending quartiles. A chi-squared test of

independence was used to test the null hypothesis that the observed distribution of prices differs from a uniform distribution. The null hypothesis was rejected at the 1 percent confidence level in all but two cases, for firms that were in the sample for a minimum of 20 or 25 weeks and that started in the second quartile.

TABLE 5
Price Quartiles Over Time

		Minimum weeks in sample				
Starting Quartile	**Ending Quartile**	**5**	**10**	**15**	**20**	**25**
1	1	63.55%	58.55%	56.65%	54.55%	58.93%
	2	20.94%	21.82%	22.66%	21.21%	21.43%
	3	13.55%	17.82%	18.72%	23.48%	19.64%
	4	1.97%	1.82%	1.97%	0.76%	0%
2	1	19.83%	25.21%	25.27%	25.69%	24.56%
	2	53.35%	44.02%	43.96%	35.78%	28.07%
	3	18.44%	20.09%	18.68%	22.02%	22.81%
	4	8.38%	10.68%	12.09%	16.51%	24.56%
3	1	11.20%	14.13%	14.88%	12.67%	11.27%
	2	22.67%	25.80%	27.44%	29.33%	32.39%
	3	49.07%	42.05%	41.40%	38.00%	43.66%
	4	17.07%	18.02%	16.28%	20.00%	12.68%
4	1	4.55%	4.56%	3.17%	3.73%	2.63%
	2	8.48%	11.20%	11.64%	12.69%	6.58%
	3	19.39%	21.58%	22.22%	25.37%	23.68%
	4	67.58%	62.66%	62.96%	58.21%	67.11%

The results strongly suggest that firms that adopt a low (first quartile) price strategy attempt to maintain their relative price ranking over time. On average, 58.45 percent of the firms that started in the first quartile ended in the first quartile and 80.06 percent of the firms ended in the first or second quartiles. The same is true of firms that adopt a high price strategy, as an average of 63.70 percent of the firms that started in the fourth quartile ended in the fourth quartile, while 86.15 percent ended in the third or fourth quartiles. Firms that enter the sample in the middle two quartiles demonstrate less 'price persistence' than firms that enter at either extreme. On average 41.1 percent and 42.84 percent of the firms that enter the sample in the second or third quartiles also exited in their respective quartiles.

CONCLUSIONS

The internet has had a profound impact on the ability of consumers to obtain information regarding both the price and quality of the products that they buy. Surveys conducted by the Pew Foundation[14] indicate that the percentage of Americans that use the internet has increased from less than 15 percent in 1995 to more than 60 percent in 2004, and that 78 percent of internet users research products online before making purchases and that 67 percent of internet users purchase goods and services online. Increased access to the internet, together with the availability of shopbots which allow consumers to obtain price quotes for a specific product from literally hundreds of sellers, have greatly reduced search costs for a large number of consumers.

Using price data for a variety of products obtained using internet shopbots, we test several hypotheses derived in the economic literature on information and optimal search, and explore the pricing behavior of firms that sell on the internet. Our results obtained using cross-sectional data indicate that the level of price dispersion is positively related to the price of the product and the number of sellers, and lower for goods that typically would be purchased several times a year. Time-series estimates, however, indicate that the price dispersion is inversely related to the price of the product and the number of sellers, and increases over time when new entrants are included in the sample; in the absence of entry the level of price dispersion decreases over time. The contradictory results obtained using the two samples illustrates the importance of studying price dispersion on the internet using both cross-sectional and time series data.

In addition, we find that controlling for handling and shipping charges and seller heterogeneity reduces the level of price dispersion by an average of 18 percent and 5 percent, respectively.

Finally, prices appear to be correlated across products and over time – low price sellers for one product generally charge low prices for all items within a given product category, and low price sellers in one time period generally are among the low price sellers in future time periods. The existence of "price persistence" across products and over time may further reduce search costs for consumers who shop online.

NOTES

We are indebted to Kerry Smith, Mark Wohar, and two referees for helpful comments. This paper is dedicated to the late Ken Koford, a friend and former colleague. Financial assistance from the Robert N. Anthony and Douglas Chair funds at Colby College is gratefully acknowledged.

1. For a discussion of shopbots see Greenwald and Kephart [1999] or Brynjolfsson and Smith [2001]. In addition, see Baye and Morgan [2001] for an analysis of the social welfare implications of information gatekeepers on the internet.
2. For example, *The Economist* (November 20,1999, 112) argued that "With perfect information about prices and products at their fingertips, consumers can quickly and easily find the best deals. In this brave new world, retailers' profit margins will be competed away, as they are all forced to price at cost."
3. When their price data are weighted by proxies for market share, they reported less price dispersion for both books and CDs sold on the internet compared to those sold in bricks and mortar stores.
4. We address this latter issue in the section Pricing Patterns across Products and over Time, where we establish that sellers of hardcover books, paperback books, and CDs generally adopt pricing policies where some firms are among the lowest priced firms for all titles while other firms are consistently among the highest.
5. The majority of previous studies in this area have measured price dispersion using the standard deviation or the coefficient of variation (the standard deviation divided by the price); we adopt the former because it is the most commonly used measure. The two measures will yield identical results as long as price is included as an explanatory variable. Ignoring the other explanatory variables, the regression model is $\ln\sigma = \alpha + \beta\ln P$ if the standard deviation is employed, and $(\ln\sigma - \ln P) = \alpha + (\beta - 1)\ln P$ if the coefficient of variation is employed. If the estimate of β is < 1 in the first equation the results imply that relative price dispersion (with respect to price) is inversely related to price. Based on the hypothesis that consumers will always purchase from the lowest price competitor, Baye, Morgan, and Sholten [2004] argue that the appropriate measure of price dispersion is the "gap", defined to be the difference between the lowest price and the average price or the lowest price and the second

lowest price, respectively. We do not employ the gap as a measure of price dispersion because the underlying hypothesis that consumers will always purchase from the firm with the lowest price is inconsistent with the data. Amazon.com is the largest on-line seller of books, but is consistently among the most expensive of the on-line book sellers, charging 5 percent and 11 percent more (on average) than Barnesandnoble.com and Borders. Com, respectively (Clay, Krishnan, Wolff, and Fernandes) [2002].

6. This finding is consistent with that reported by Clay et al. [2001].
7. Bailey [1998] reports that internet retailers change prices more than twice as often as traditional retailers.
8. Carlton and Chevalier [2001] report that shipping costs appear to be positively related to the price charged for fragrances by internet sellers. Clay, Krishnan, Wolff, and Fernandes [2002] report that at least two of the firms in their sample followed a low price, high shipping cost strategy.
9. The regression results are generally robust with respect to the attribute chosen. Pan, Ratchford, and Shankar [2002] and [2003] also report that the various measures of seller quality are highly correlated; they employ factor analysis to choose the measures included in their regression equations.
10. Stigler's argument may be applied to a firm's pricing policy across products as well as over time.
11. Kendall's W may be viewed as an average of the Spearman rank correlation coefficients for all items within a product class. For more details, see Siegel [1956], 229-239.
12. For example, if one firm were represented in the sample for the first twelve weeks and a second firm for the last thirteen weeks, then the Kendall test would delete the data for all twenty-five weeks.
13. The price rankings are computed using only the firms included in the reduced sample.
14. The Pew Foundation has an ongoing research project dealing with the Internet and American Life. The survey results may be found at http://www.pewinternet.org/trends/InternetAdoption.jpg and http://www.pewinternet.org/trends/Internet_Activities_8.05.05.htm.

REFERENCES

Bailey, J. P. *Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce.* Ph.D. Thesis, Technology, Management and Policy, MIT, 1998.

Baye, M. and Morgan, J. Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets. *American Economic Review,* June 2001, 454-474.

_____**.** Price Dispersion in the Lab and on the Internet: Theory and Evidence. *Rand Journal of Economics*, Autumn 2004, 448-466.

Baye, M., Morgan, J. and Scholten, P. The Value of Information in an Online Consumers Electronic Market. *Journal of Public Policy and Marketing*, Spring 2003, 17-25.

_____**.** Price Dispersion in the Small and Large: Evidence From an Internet Price Comparison Site. *Journal of Industrial Economics*, December 2004, 463-496.

_____**.** Persistent Price Dispersion in Online Markets, in *The New Economy and Beyond: Past, Present and Future,* edited by D. W. Jansen and Edward Elgar. Northampton MA: Edward Elgar Press, 2006.

Brown, J. R. and Goolsbee A. Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy*, June 2002, 481-507.

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

_____. The Great Equalizer? Consumer Choice Behavior at Internet Shopbots, MIT Sloan School Working Sloan Working Paper 4208-01, October 2001.

Carlson, J. and McAfee, R. P. Discrete Equilibrium Price Dispersion. *Journal of Political Economy*, June 1983, 480-93.

Carlton, D. W. and Chevalier, J.A. Free Riding and Sales Strategies for the Internet. *Journal of Industrial Economics*, December 2001, 441-461.

Chen, P. and Hitt, L. M. Understanding Price Dispersion in Internet-Enabled Markets. Working Paper, Carnegie Mellon University, Pittsburgh, PA, 2003.

Clay, K., Krishnan, R. and Wolff, E. Prices and Price Dispersion on the Web: Evidence From The Online Book Industry, *Journal of Industrial Economics*, December 2001, 521-539.

Clay, K., Krishnan, R., Wolff, E. and Fernandes, D. Retail Strategies on the Web: Price and Non-Price Competition in the Online Book Industry, *Journal of Industrial Economics*, September 2002, 351-367.

Dahlby, B. and West, D. S. Price Dispersion in an Automobile Insurance Markets, *Journal of Political Economy*, April 1986, 418-38.

Dana, J. D., Learning in an Equilibrium Search Model, *International Economic Review*, August 1994, 745-771.

Friberg, R., Ganslandt, M. and Sandstrom, **M**. E-Commerce and Prices – Theory and Evidence, Economic Research Institute, Stockholm School of Economics Working Paper in Economics and Finance No. 389, September 2000.

Greenwald, A. R. and Kephart, J.O. Shopbots and Pricebots, *The Proceedings of International Joint Conference on Artificial Intelligence*, 1999.

Janssen, M. C. W. and Moraga, J. L. Pricing, Consumer Search and the Size of Internet Markets, Tinbergen Institute Discussion Paper TI 2000-0042/1, May, 2000.

Pan, X., Ratchford, B. T., and Shankar, V. Can Price Dispersion in Online Markets be Explained by Differences in E-tailer Quality? Journal of the Academy of Market Science, October 2002, 433-445.

_____. Why Aren't the Prices of the Same Item at Me.com and You.com the Same?: Drivers of Price Dispersion Among E-Tailers Working paper, University of Maryland, College Park, MD, July 2003.

Pratt, J. W., Wise D. A., and Zeckhauser, R. Price Differences in Almost Competitive Markets, *Quarterly Journal of Economics*, May, 1979, 189-211.

Ratchford, B. T., Pan, X. and Shankar, V. On the Efficiency of Internet Markets for Consumer Goods, *Journal of Public Policy and Marketing*, Spring 2003, 4-16.

Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, New York, 1956.

Sorensen, Alan T., Equilibrium Price Dispersion in Retail Markets for Prescription Drugs, *Journal of Political Economy*, August, 2000, 833-50.

Stahl, D. O., Oligopolistic Pricing with Sequential Consumer Search, *American Economic Review,* September 1989, 700-712.

Stigler, G., The Economics of Information, *Journal of Political Economy*, June 1961, 213-225.

Varian, H. R., A Model of Sales, *American Economic Review*, September, 1980, 651-659.

White, H., A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, May,1980, 817-838.