



Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com

JUDITH CHEVALIER

*Yale School of Management, 135 Prospect Street, New Haven, CT 06520
and National Bureau of Economic Research
E-mail: judith.chevalier@yale.edu*

AUSTAN GOOLSBEE

*University of Chicago Graduate School of Business, 1101 E. 58th Street, Chicago, IL 60637,
National Bureau of Economic Research and American Bar Foundation
E-mail: goolsbee@gsb.uchicago.edu*

Abstract. Despite the interest in measuring price sensitivity of online consumers, most academic work on Internet commerce is hindered by a lack of data on quantity. In this paper we use publicly available data on the sales ranks of about 20,000 books to derive quantity proxies at the two leading online booksellers. Matching this information to prices, we can directly estimate the elasticities of demand facing both merchants as well as create a price index for online books. The results show significant price sensitivity at both merchants but demand at BarnesandNoble.com is much more price-elastic than is demand at Amazon.com. The data also allow us to estimate the magnitude of bias in the CPI due to the rise of Internet sales.

Key words. internet, retail, perfect competition, price indices, elasticity

JEL Classification: D41, D43, L1, L81, L86, M31

In the earliest days of Internet commerce, many economists and media observers predicted that competition among Internet retailers would quickly resemble perfect competition.¹ After all, the Internet already reduces search costs relative to visiting physical stores, and shopbots and comparison sites could be expected to lower search costs still further.

Two strands of research have addressed the question of price competition on the Internet. The first set of papers examines patterns of prices for homogeneous goods. Researchers have examined the degree of price dispersion amongst Internet retailers, as well as between Internet retailers and bricks-and-mortar retailers. The general consensus of these papers is that price dispersion amongst Internet retailers is large, and that online retailers charge prices that are either modestly lower or actually

¹ See, for example, Kuttner (1998).

higher than their offline counterparts.² These results seem incompatible with the idea that the Internet has completely eliminated consumer search costs. An important advantage of this strand of research is that these studies require only publicly available price data. However, a concern with these findings is that, while relatively high prices are posted at some Internet sites, few or no transactions may be taking place at those relatively high prices. Without quantity data, it is impossible to know.

A second strand of research attempts more direct measures of consumer price sensitivity.³ The general consensus from this work seems to be that Internet markets do seem competitive in the sense that demand for a seller appears to be quite elastic to the seller's own price or to competitors prices. One important drawback of this research is that all of the papers rely on proprietary information on firm sales or consumer buying patterns. In particular, much of this work focuses on consumers who use Internet shopbots, a group that may not be representative of overall Internet shoppers.⁴ In general, there has been little overlap in the industries studied by the two approaches.

We examine online books, in part because this is the most-studied Internet retail category, but also because it is one of the largest online sales categories. We develop a method to estimate directly the own- and cross-price elasticities of demand at Amazon.com and BarnesandNoble.com (hereafter, BN.com). We also compute a Fisher-ideal price index for online books. To do these things we need only two sources of data: publicly available information on prices and sales ranks at the two leading sites and data from simple experiments which anyone can conduct for less than \$50.

Our results show several things about prices and competition in the online book industry. First, having sales data matters for the results. The prices of online books, for example, look dramatically different when books are weighted by sales compared to when all books are weighted equally (as assumed in the conventional literature). Also, it is clear that online inflation behaves quite differently in this period than does the CPI for recreational books. Indeed, our best estimates suggest that the CPI misstates the true inflation rate by almost a factor of three in one part of our sample

0 Work by Lee (1997) for cars and Bailey (1998) for books, CDs, and software suggest that prices were actually higher online than in retail stores. More recent work by Brynjolfsson and Smith (2000) for books and CDs and by Clay et al. (2000) for books has found prices the same or lower online but that online price dispersion is quite high, perhaps greater than in retail stores. Carlton and Chevalier (2001) show, among other things, the existence of price dispersion among online fragrance retailers.

3 Goolsbee (2000, 2001) finds a large cross-price elasticity of online retail and online computers with respect to physical retail prices. Ellison and Ellison (2001) find large elasticities for computer memory and motherboards from data on a private computer parts retailer. Brown and Goolsbee (2002) and Scott Morton et al. (2001) examine the impact of Internet shopbots on prices of life insurance and for cars respectively and find that the Internet leads to significantly lower prices. Smith and Brynjolfsson (2001) examine customer behavior at a book price comparison site but find that brand still matters for consumers' click through probabilities.

4 See White (2000) for a discussion of the usage of Internet shopbots.

and gets the sign wrong in another. Second, we show that there is significant price sensitivity for online book purchases at both sites. The demand at BN.com, however, is much more price sensitive, both to its own and to the rival's price, than is demand at Amazon.com. Third, looking across different time periods, our results show that using measured price dispersion to infer the degree of price competition, as is commonly done in the literature, can be misleading.

The question of how pricing impacts consumer purchasing online is interesting in and of itself, but also has implications for public policy questions. In this paper, we discuss one such application, measuring the potential magnitude of bias in the consumer price index arising from underweighting Internet commerce. The potential importance of bias relating the changing pattern of outlets shopped by consumers has been highlighted in other literature (see Schultze and Macki, 2002; Boskin et al., 1996; Reinsdorf, 1993) but this is one of the first pieces of direct evidence on the subject and the only one relating to the rise of the Internet.

Finally, this project provides a contribution to the growing field of web mining (see Kosala and Blockeel, 2000, for a survey). Our particular technique for turning ranks into sales is specifically applicable only in situations in which sales rank data are available while quantity data are not. However, more generally, one could use very similar distributional assumptions and data-gathering techniques to transform other web-gathered measures, such as page view data, web counter data, or site visit data into sales proxies.

The paper proceeds as follows. Section 1 provides the background and describes the data. Section 2 presents the methodology for translating sales ranks into sales quantities. Section 3 presents price indices for Amazon.com and BN.com and assesses the impact of price movements at Amazon.com and BN.com on CPI bias. Section 4 provides evidence on the demand elasticities. Section 5 briefly describes some robustness checks. Section 6 concludes.

1. Background and data

Amazon.com began selling books online in 1995, one of the first electronic commerce firms. By 1999, books were the second largest retail segment (after computers) sold over the Internet (BCG, 2000). Online book sales grew from essentially nothing in 1995 to more than \$2 billion in 2000 (Forrester, 2001). Today such sales make up between 7.5% and 10% of total book sales in the United States (American Booksellers Association, 2002; Cader, 2001). Within the online bookstore industry, the two dominant players are Amazon.com and BN.com. According to Net Ratings, these two firms account for more than 85% of online book sales. Recent accounting data shows that Amazon.com's book, music and DVD sales are approximately four times that of BN.com.

More detail about the business operations of the online merchants can be found in Rayport (1998). For purposes of this paper, what is relevant is that online bookstores tend to have much larger selection of titles than even the largest physical

bookstores. A large superstore might have as many as 150,000 titles whereas Amazon.com and BN.com claim to have millions of titles available (although for books outside of the top 200,000, this may involve waiting two weeks or more to actually receive the book).

A customer visiting one of the sites and looking for a book would typically face a screen giving the price of the book, the relative sales ranking at the site, information on the shipping time/availability, a brief description of the book, customer reviews of the book and a listing of other books and authors that are popular among people interested in the book, and the price for a used version of the book (if available).

We collected data during three different weeks in 2001 on about 20,000 different books from the websites of Amazon.com and BN.com. We did this by ISBN number. Since a tiny fraction of books in print account for most book sales, building a large and representative sample is not easy. To get books representative of different parts of the sales distribution, we combined ISBN numbers from three sources. First, we included all books that appeared on any *Publishers' Weekly* best-seller list from 1996 to March 2000 (predating our sample). Second, we include all books that were searched for at Dealtime.com from August 25 to November 1 1999 as compiled in Smith and Brynjolfsson (2001).⁵ Third, we took a random sample of about 3000 books (Book in Print, 2000). In total, these three methodologies give us approximately 26,000 ISBN numbers of which typically about 18,000 had price and rank data at Amazon.com and about 13,000 had price and rank data at BN.com. The difference arises because BN.com does not report sales ranks for books with rankings greater than about 630,000 whereas Amazon.com's are not censored (and go to over 2,000,000). We will address this asymmetric censoring of rankings in our results below.

Our three samples were taken during the weeks of April, June and August of 2001. During this period there were major price changes by both sellers. We do not look at price changes over very short time horizons because of the way the ranks are updated at the sites. Amazon.com claims that for books in the top 10,000 ranks, the rankings are based on the last 24-hours and updated hourly. For books ranked 10,001–100,000, the ranks are updated once per day. For books ranked greater than 100,000, the sales ranks are updated once per month (Amazon.com, 2000). Many hundreds of thousands of books, however, have a rank but almost certainly have less than one sale per month. Italic (2001) claims that for these rarely purchased books, Amazon.com bases the rank the total sales since Amazon.com's inception. BN.com claims to update all the rankings daily (BN.com, 2000).⁶

⁵ We thank Michael Smith and Erik Brynjolfsson for providing this list to us.

⁶ Since BN.com provides rankings on tens of thousands of books that average far less than one sale per day, this statement cannot be completely accurate. They would not provide us any more detail in their ranking system (despite repeated requests).

In the first period of our sample, taken during the week of April 13 2001, prices had been quite stable for some time. The general price structure at Amazon.com and BN.com was to discount hardback books at 20% off their retail price, paperback books at 10% off, *New York Times* bestsellers at 40% off and textbooks at no discount (some other types of books were also sold with no discounts and there are periodic editor picks and the like that receive further discounts). The sites do differ in their classifications of some of the books and Amazon.com tends to use *New York Times* bestseller lists with a lag whereas BN.com does not.

Starting June 20 2001, Amazon.com conducted a two-week pricing experiment in which it raised the prices of many of its books. Our second data collection occurred during the week of June 23 2001. Amazon.com announced the launch of free shipping for all buyers purchasing more than two books while simultaneously increasing overall prices rather significantly.⁷ During this period, they eliminated all discounts for most paperback books, maintained no discount for textbooks, and reduced the discount on hardback books to 10%. BN.com generally maintained their previous pricing structure.

The pricing regime of June did not last long. On July 3, BN.com launched free shipping with the purchase of two items. At that time, BN.com vice chairman Steve Riggio contrasted the BN.com strategy to Amazon.com's by noting "we're offering free shipping without changing our prices or making any fine-print exceptions." On July 4, Amazon.com removed the free shipping offer and changed prices again. The company claimed that the two-week price change was merely an experiment and that it was intended to be short-lived.

In the third period of our sample (conducted during the week of August 3 2001), Amazon.com had reinstated the 20% discount but now applied it only to books over \$20. Books under \$20 generally received no discount, nor did textbooks. In this new period, BN.com's policy was not explicitly stated but they appeared to move away from the standard discounts of 10% and 20% for paper and hardback books.

Importantly for our estimation, the pricing at these sites is set at a general level. That is, broad categories of books all receive the same discount off of the manufacturer's suggested retail price. Individual book pricing appears to be done only for a very small number of editor's picks. The sales of a particular book (relative to the book's broad category) does not seem to impact the book's pricing. During this time period, price differences between the sites mostly reflect differences in the prices charged for a particular category of book or differences in the categorization scheme (for example, whether or not the *Chicago Manual of Style* is classified as a textbook).

⁷ Amazon.com made the following statement during this time: "We've also changed our pricing on some books, CDs, DVDs, and videos: for some products prices have stayed the same, for some products prices are lower, and for some products we've reduced our discounts." (www.internetnews.com, July 2 2001). However, our observation from the data is that prices mainly increased.

2. Computing sales quantities

Our basic approach is to translate the observed sales ranking of each book into a measure of quantity. To do so, we need to know the probability distribution of book sales. A standard distributional assumption for this type of rank data is a Pareto distribution (i.e., a power law).⁸ In the Pareto distribution, the probability that an observation, s , exceeds some level, S , is an exponential function

$$\Pr(s > S) = \left(\frac{k}{S}\right)^\theta, \quad (1)$$

where k and θ are the parameters of the distribution. The most important parameter is θ , the shape parameter that indicates the relative frequency of large observations. If θ is 2, for example, the probability of an event decreases in the square of the size. With a value of 1, it decreases linearly.

For any given book, the number of books that have greater sales than that book is just the book's rank $- 1$. Therefore, the fraction of all books that have sales greater than a particular book is just $(\text{Rank} - 1)/(\text{Total Number of books})$. If there are a sufficient number of books to eliminate discreteness problems, then we can replace equation (1) with:

$$\frac{(\text{Rank} - 1)}{\text{Total number of books}} = \left(\frac{k}{S}\right)^\theta. \quad (2)$$

Taking logs, we can translate sales into ranks according to

$$\ln(\text{Rank} - 1) = c - \theta \ln(\text{Sales}). \quad (3)$$

The constant c is a function of θ , k , and the total number of books for sale. In what follows, we will constrain θ to be the same for both Amazon.com and BN.com, but we will not constrain c to be the same across sites.

Evidence that the Pareto distribution fits well for books can be found using the weekly *Wall Street Journal* book sales index which, unlike other bestseller lists, gives an index of the actual quantity sold. This index is constructed by surveying Amazon.com, BN.com, and several large brick and mortar book chains. Using the data from April to August 2001, we regress $\log(\text{rank} - 1)$ on \log sales for each book-

⁸ For discussions on the use of power law distributions to describe rank data, see Pareto (1897), Quandt (1964), and Mandelbrot (1988). It should be noted that the well-known "Zipf's law" (Zipf, 1949) is a special case of Pareto's law. More details on the Pareto and its application can be found in Johnson and Kotz (1970) or Goolsbee (1999).

week observation in the data set, as well as weekly dummies. The regression specification fits very well. The R -squared of this regression is 0.94. The estimated value of θ is 1.49.

We could use this estimate of θ to translate sales ranks into quantities in our main sample, but sales online may have a different distribution than sales in stores. We are able to get several independent estimates of θ strictly for online books (as described below) and they are all quite close.

The first estimate of θ comes from a single publisher who was monitoring the sales of their own books and competitor's books on Amazon.com. We discovered that this publisher was using a methodology very similar to our own. This publisher had performed a non-linear least squares regression of the form $\text{Rank} = A \times (\text{Sales})$ using their own proprietary data on the sales of their own books at Amazon.com. Of course, this methodology should produce the same results as a linear regression of $\ln(\text{Rank})$ on $\ln(\text{Sales})$. Given that most of this publisher's books had ranks of 100 or greater, we are not particularly concerned that the publisher's methodology did not subtract one from the ranks, as the Pareto formula requires. This was done for the subsample of this publisher's books that had sales ranks at Amazon.com in the top 15,000 over the course of one week. They conducted two regressions giving estimates of θ of 0.9 and 1.3. They did not provide us with standard errors on either estimate.⁹

For the second estimate, we conducted our own experiment. A publisher who would not give us direct information on rankings and sales of their books did tell us of a title they had with steady sales at Amazon.com of 14 copies per week (i.e., about two copies per day). We observed this book to be ranked 14,468. We then purchased six copies of the book in a 10-minute period and observed its rank rise to 2854. Assuming two sales in a day corresponded to the first ranking and eight sales in a day corresponded to the second ranking, we can solve for the implied Pareto shape parameter, θ . In this case it is equal to 1.17.

Third, an author, Gene Weingarten did a similar experiment to ours with his own book (see Weingarten, 2001). According to the author, his new book's ranking was 1,484,129. Purchases of 20 copies in an hour sent the book to rank 1297. Purchases of another five copies moved the book to rank 1025. Assuming that the daily sales at a rank of 1,484,129 is very close to zero, this implies a θ of 1.05.

Finally, Poynter (2000), gives an estimate of actual sales in seven different rank ranges (e.g., ranks 450 to 750 average 90 sales per week). Taking the mid-point of his ranges and regressing the log rank on the log sales yields an estimate of θ of 1.199 (with a standard error of 0.102 and an R^2 of 0.97).

Thus all of these experiments suggest fairly consistent estimates of θ in the relatively tight range of 0.9 to 1.3 and can be used to translate ranks into sales. We will use 1.2 as the basic estimate.

⁹ See Schnapp and Allwine (2001), for a description of the methodology.

3. Price indices and price dispersion

Given this estimate of the shape parameter, we can compute the implied sales for any book that has a sales rank. Using these sales weights, we can also compute a price index to compare prices across sites within a time period or across time within a site. Since the BN.com books are censored at ranks of approximately 600,000, we restrict the sample here to books with data on ranks in all periods at both sites.¹⁰ The price indices will allow us to determine, with proper weighting, the rate of inflation at the online book stores as well as to know which of the sites is more expensive.

We start, in Table 1, by showing that it can be somewhat misleading to use unweighted prices. Previous work has, by necessity, not had sales weights and thus has calculated equally weighted price indices. The first two panels show that the

Table 1. Prices and price indices.

	BN.com price	Amazon.com price
Books equally weighted (% discount)		
April 14 2001	\$24.23 (11.3%)	\$23.58 (15.4%)
June 23 2001	\$24.39 (10.9%)	\$26.75 (2.1%)
August 3 2001	\$24.51 (10.6%)	\$22.26 (19.2%)
Books equally weighted, including shipping		
April 14 2001	\$25.22	\$24.57
June 23 2001	\$25.38	\$26.75
August 3 2001	\$24.51	\$23.25
Books weighted by sales (% discount)		
April 14 2001	\$13.09 (34.9%)	\$14.32 (26.1%)
June 23 2001	\$13.24 (34.0%)	\$15.72 (18.4%)
August 3 2001	\$13.12 (33.4%)	\$15.25 (29.7%)
Books weighted by sales, including shipping		
April 14 2001	\$14.08	\$15.31
June 23 2001	\$14.23	\$15.72
August 3 2001	\$13.12	\$16.24
Price index within period, across site		
April 14 2001	100.0	96.7
June 23 2001	100.0	103.4
August 3 2001	100.0	95.3
Price index within site, over time		
April 14 2001	100.0	100.0
June 23 2001	100.7	109.5
August 3 2001	98.6	105.3

Authors' calculations as described in the text. When shipping is included, it is marginal shipping assuming the customer is buying this as their second book. Sales weights in the second panel are current period sales estimated using a Pareto shape parameter of 1.2 as described in the text.

10 Given that the previous section demonstrated that sales are dropping almost linearly with the sales rank, there is little impact of the cutoff rule on the results.

unweighted prices are notably different from the weighted (where the sales weights here are just the actual sales of the given book in the period as estimated using a Pareto shape parameter θ of 1.2 and normalizing the sales of the highest selling book to 1). Sales of inexpensive books are much greater than the sales of expensive books, leading the sales-weighted average prices to be less than half of the raw averages.¹¹ The percent discount from the manufacturer's list price is given in parentheses and shows the same major difference between the weighted and unweighted data.

To construct a proper price index we use the sales quantities at the two sites and compute a Fisher ideal index in each of the three periods (with BN.com providing the base of 100.0 in each). This is reported in the middle panel of Table 1. In the first period, Amazon.com has lower prices than BN.com. In the second period, with Amazon.com's price increase experiment, Amazon.com's prices are about 3.4% higher than BN.com's. In the week of August 3rd, Amazon.com's prices are again lower than at BN.com (note that these indices show only prices at Amazon.com relative to BN.com at each point in time, not the prices within Amazon.com across time).

To examine the extent of inflation in the same site over time, the bottom panel of the table creates a Fisher Ideal, chain-weighted price index where the base period at each site is 100.0 in the week of April 14th (note that the price indices here are not comparable across sites, only across time for a given site). The indices show modest inflation at BN.com from April to June followed by noticeable deflation from June to August. Prices at Amazon.com behaved even more dramatically in this time period, showing almost 10% inflation in a period of only two months followed by significant deflation over the next month and a half.

These dramatic price movements for online books could have potentially important consequences for consumers. However, online book vendors are not sampled in constructing the consumer price index (CPI) for recreational books. As outlined by the Bureau of Labor Statistics (Cage, 1996), the CPI relies on the Point

Table 2. Inflation using different price indices (in percent).

Period	CPI	BN.com	Amazon.com	"True" inflation
April 14 2001 → June 23 2001	0.4	0.7	9.5	1.0
June 23 2001 → August 3 2001	0.1	-2.1	-3.8	-0.2
April 14 2001 → August 3 2001	0.5	-1.4	5.3	0.8

The inflation calculation uses the prices without shipping costs.

11 These prices do not include shipping charges. As we discuss later, it is not obvious how to best include shipping prices since the shipping charge schedule is non-linear and there are multiple shipping options available at each site. Here, the inclusion of any shipping charges would raise BN.com and Amazon.com's prices symmetrically in period 1, as the two sites had identical shipping price schedules at that time. In period 2, Amazon.com's price would fall relative to BN.com's price, as Amazon.com was offering free standard shipping with the purchase of two items. For a benchmark, the marginal price of shipping a third book at Amazon.com would be zero, versus 0.99 cents at BN.com. In period 3, BN.com's relative price would be lower it was offering free standard shipping with the purchase of two items, while Amazon.com had reverted to its previous shipping schedule.

Table 3. Price dispersion between Amazon.com and BN.com.

	Price dispersion
April 14 2001	0.081 (0.104)
June 23 2001	0.095 (0.108)
August 3 2001	0.123 (0.130)

Authors' calculations. Price dispersion calculated as $2(|\text{Price}_{\text{AMZN}} - \text{Price}_{\text{BN}}|)/(\text{Price}_{\text{AMZN}} + \text{Price}_{\text{BN}})$. Standard deviation in parentheses.

of Purchase Survey to determine what types of retail outlets the BLS should survey when including price quotes in the CPI. These weights enter the sampling only with a lag of several years. In the case of books, the weights in the CPI during our period are based on purchase patterns from 1995–1998. At the start of that time, online books sales were virtually nonexistent. Even in 1998, online books sales were significantly smaller than at the time of our sample. Since today online sales account for close to 10% of book purchases, this could mean a serious retail outlet bias.

In Table 2, we show the inflation rate at the two online sites over the period and the inflation rate as given by the official CPI for recreational books. In the final column of the table, we approximate the true inflation rate assuming that the CPI accurately reflects the behavior of bricks and mortar booksellers and that it accurately reflects the behavior of online booksellers other than Amazon.com and BN.com. We then recalculate a CPI for recreation books, giving our estimated price series for Amazon.com and BN.com but giving them a share weight of 8% of total books (with Amazon.com having 75% of that). As the table indicates, the bias caused by neglecting the Internet merchants is extreme. Our best estimate of the inflation rate suggests that the CPI's inflation rate for retail books is mistaken by more than a factor of two in the first period and gets the wrong sign in the second period. As the first (to our knowledge) micro estimate of the bias in the CPI arising from the Internet, it suggests further examination.¹²

In addition to a price index, we can also compute the degree of price dispersion across sites. Sorenson (2000), Brynjolfsson and Smith (2000), and Clay et al. (2000) use the dispersion across Internet retailers as a measure of the imperfection of market competition. In Table 3, we calculate the degree of price dispersion across the sites for the three time periods. This measure is calculated as:

$$\text{Price dispersion} = \frac{1}{N} \sum_{i=1}^N \frac{|\text{Price}_{i,\text{AMZN}} - \text{Price}_{i,\text{BN}}|}{(\text{Price}_{i,\text{AMZN}} + \text{Price}_{i,\text{BN}})/2}. \quad (4)$$

¹² Notice that our calculations assume that the fraction of consumers purchasing books online is independent of the ratio of online to bricks and mortar prices. Obviously, if the ratio of online to offline sales responds to the price ratios, the outlet substitution bias that we identify could be magnified.

That is, price dispersion measures the extent to which BN.com and Amazon.com are charging different prices from one another in each period. In the framework of the existing literature, which interprets price dispersion as evidence of market power, the results in Table 3 would be interpreted as suggesting that the level of competition between the two sites has been diminishing over time, as price dispersion is steadily rising. Using our measures of elasticities, we will be able to estimate the amount of competition in the different periods and thus check the validity of the existing approach. Previewing the results, we do not find any evidence that consumer price sensitivity has fallen over time.

The data on price differences across sites and across time also motivates such empirical analysis since it means there is likely to be considerable variation across sites for price of the same books. Another way to think about the price variation is to note that, pooling all the periods together, about 36% of the books have identical prices across the two sites. Amazon.com's prices are higher for about 28% of the books and lower for about 35%. Of the two-thirds of books where prices differ across site, about nine out of ten of them have price differences of more than 5%.

4. Estimating the demand system

4.1. Empirical approaches

Given the price variation across sites and across time and the measures of quantity each period coming from the sales ranks, we can consider the price sensitivity of online book sales. We start with each cross-section separately, and ask, essentially, whether relative sales across sites are lower when the relative price is higher.

Calling the total sales of book b at site s during week t , Q_{bt}^s , and log sales q_{bt}^s , we assume that the log sales of a book depends on book and site dummies, the characteristics of the book at a site, denoted x , including things like the shipping time and the customer reviews, and the log of the price at the site, p_{bt}^s , and at the competitor site, p_{bt}^{-s} according to

$$q_{bt}^s = f_b + f^s + \beta^s p_{bt}^s + \alpha^s p_{bt}^{-s} + \Gamma' x_{bt}^s + \varepsilon_{bt}^s. \quad (5)$$

With cross-sectional book data across sites, we can estimate a relative elasticity of substitution by estimating the relative demand as a function of relative prices

$$(q_{bt}^s - q_{bt}^{-s}) = (f^s - f^{-s}) + (\beta^s - \alpha^{-s})(p_{bt}^s) + (\beta^{-s} - \alpha^s)p_{bt}^{-s} + \Gamma'(x_{bt}^s - x_{bt}^{-s}) + \eta_{bt}. \quad (6)$$

As the equation makes clear, the coefficients on log price are not the true elasticities but rather a combination of the own and cross-price elasticities of demand because a change in the price at one site affects relative demand in two ways. One is by reducing demand via the negative own price elasticity. The other is by raising the

demand at the other site via the cross-price elasticity. Notice that, in this estimation, the estimated constant term is equal to the difference in the constant terms across the two sites. Thus, we are implicitly allowing the Pareto k parameter and the total number of books to differ across the two sites. However, we are constraining the Pareto k parameter and the total number of books to be the same within site, across time.

With more than one time period and with price changes over time, we can estimate the own- and cross-price elasticities directly. Differencing the data across time, we see that

$$(q_{bt}^s - q_{bt-1}^s) = (f_t - f_{t-1}) + (\beta^s)(p_{bt}^s - p_{bt-1}^s) - (\alpha^s)(p_{bt}^{-s} - p_{bt-1}^{-s}) + \Gamma'(x_{bt}^s - x_{bt-1}^s) + \omega_{bt}, \quad (7)$$

which gives the demand coefficients α and β separately.¹³

To translate this model into one that we can use rank data for requires us only to substitute the log sales rank for the log quantity in equation (4) according to the Pareto relationship in equation (1).¹⁴

$$(r_{bt}^s - r_{bt-1}^s) = \theta(f_t - f_{t-1}) + \theta(\beta^s)(p_{bt}^s - p_{bt-1}^s) - \theta(\alpha^s)(p_{bt}^{-s} - p_{bt-1}^{-s}) + \theta\Gamma'(x_{bt}^s - x_{bt-1}^s) + \theta\omega_{bt}. \quad (8)$$

In other words, estimating the equations using log ranks, r , rather than actual quantities, yields the correct elasticity but scaled up by the Pareto shape parameter, θ , which we estimated above.

5. Results

5.1. Cross-sectional results

Before estimating the parametric model, we present suggestive evidence about price sensitivity that does not rely on the Pareto assumption. Table 4 presents the mean difference in log ranks for books where the prices are lower at Amazon.com than at BN.com, the same as at BN.com, and greater than at BN.com. In the pooled sample and in every period individually, when the relative price is lower at Amazon.com, the sales ranks also tend to be lower (meaning greater sales). In August of 2001, for example, among books whose prices were higher at Amazon.com than at BN.com,

13 Of course, it is equivalent here or in (3) to include book and time period or book and site (respectively) fixed effects rather than differencing the data.

14 For simplicity we refer to the dependent variable as the log rank rather than being more precise and calling it the log of the sales rank minus one which we use in the empirical work.

Table 4. Relative sales ranks across sites as a function of relative prices compares ranks as a function of the relative prices of observations at Amazon.com and BN.com.

	$P_{AMZN} > P_{BN}$	$P_{AMZN} = P_{BN}$	$P_{AMZN} < P_{BN}$
Number of books	10,089	13,297	12,602
Avg. $\ln(\text{Rank}_{AMZN}) - \ln(\text{Rank}_{BN})$			
<i>Pooled:</i>	-0.017	-0.225	-0.627
<i>By week:</i>			
April 14 2001	0.232	-0.142	-0.662
June 23 2001	-0.151	-0.569	-0.580
August 3 2001	0.313	-0.008	-0.607

Amazon.com's sales ranks averaged about 31% higher than at BN.com. For books whose prices were the same at the two sites, Amazon.com ranks were less than 1% lower than BN.com's. Among books whose prices were lower at Amazon.com, the sales ranks at Amazon.com were about 61% lower than at BN.com. It is important for this analysis that the universe of available books is the same at both sites (which we believe approximately true). We do find that our results are robust to restricting the sample to books which rank above 75,000 at both sites. As both sites purport to have relatively comprehensive catalogs, it seems unlikely that, in the range of relatively popular books, the sites differ much in their selection.

In column 1 of Table 5, we present a probit specification. The dependent variable takes the value of one if the sales rank at Amazon.com is higher than the sales rank at BN.com (note: higher ranks correspond to lower sales). The independent variables are the relative price at Amazon.com (P^A/P^B) and time dummies. The coefficient on the relative price shows a large and statistically significant positive coefficient. That is, when Amazon.com has relatively higher prices, it has relatively higher ranks (lower sales).

To compute a price elasticity, however, we invoke the Pareto assumption. Because the sales rank data are censored at BN.com, we estimate equation (3) using the trimmed least absolute deviations (LAD) panel estimator of Honore (1992). The Trimmed LAD is used due to the difficulty with and inconsistency of a Tobit estimator with a large number of unit fixed effects (but a small number of time periods). Because we don't know the exact censoring point, we used the largest observed rank at BN.com in the sample. We treat Amazon.com observations as uncensored.

The results are reported in columns (2)–(4) of the table. The dependent variable is the sales rank. The explanatory variables included in the regression are the price, a site dummy, availability dummies (i.e., ships in 24 hours, 2–3 days, and so on) and dummies for each individual book title. Including dummies for individual book titles is crucial. This ensures that our identification derives from the differences in prices of the same book across sites.

The top panel uses prices without shipping charges. As there are numerous shipping choices (ground, air, priority, FedEx) and the prices depend on how many

Table 5. Cross-sectional evidence on price sensitivity of demand.

	(1) Probit Pooled	(2) Trimmed LAD April 14 2001	(3) Trimmed LAD June 23 2001	(4) Trimmed LAD August 3 2001
$\ln(P_{AMZN}/P_{BN})$	2.573 (0.055)			
$\ln(P)$		2.135 (0.080)	2.550 (0.105)	2.124 (0.068)
Shipping dummies	No	Yes	Yes	Yes
Amazon.com dummies	No	Yes	Yes	Yes
Time dummies	Yes	No	No	No
ISBN dummies	No	Yes	Yes	Yes
n	34,156	33,650	27,094	30,638
$\ln(P + \text{Shipping})$		2.276 (0.085)	1.919 (0.095)	2.458 (0.069)
Shipping dummies		Yes	Yes	Yes
Amazon.com dummies		Yes	Yes	Yes
Time dummies		No	No	No
ISBN dummies		Yes	Yes	Yes
n		33,650	27,094	30,638

Dependent variable in column 1 is the $\{0,1\}$ variable of whether the rank is higher at Amazon.com than at BN.com. Dependent variable in columns 2–4 is the log of the rank. This is censored as described in the text. Standard errors are in parentheses.

books one orders according to a two-part tariff, it isn't clear what shipping price to use. The lower panel uses prices that include the incremental shipping charge if a person was adding this as a second book to their order. This was 99 cents at both sites in period 1, zero at Amazon.com and 99 cents and BN.com in period 2, and then zero at BN.com and 99 cents at Amazon.com in period 3. The results do not differ much between the two (nor with other choices of the shipping price), so we will not include shipping in the subsequent regressions.

All price coefficients are highly significant and in the range of 2.1 to 2.5. With our mean estimate of the Pareto shape parameter, this indicates an elasticity -2.5 and -3 . In words, this says that a 1% increase in the price of a book at site A relative to site B reduces sales at A by 2.5 to 3% relative to sales of that book at site B.

5.2. Panel results

The relative price elasticity above is notably large, suggesting the importance of competition. To break out the own- and cross-price components of this number, however, requires variation in prices across time. To get such variation, we use pairs of time periods for each site. One reason to do this is that we can then use trimmed LAD estimation to allow for the censoring problem with the BN.com

Table 6. Two period panel estimates of online book demand system.

Dep Var.:	(1)	(2)	(3)
ln(Rank)	t_2, t_1	t_3, t_2	t_3, t_1
<i>BN.com (trimmed LAD)</i>			
ln(P_{own})	4.396 (0.182)	2.985 (0.128)	2.894 (0.128)
ln(P_{cross})	-3.825 (0.181)	-2.403 (0.128)	-2.696 (0.114)
Shipping dummies	Yes	Yes	Yes
Time \times age dummies	Yes	Yes	Yes
ISBN dummies	Yes	Yes	Yes
n	24,738	24,738	24,738
<i>Amazon.com (OLS)</i>			
ln(P_{own})	0.262 (0.032)	0.256 (0.048)	0.371 (0.050)
ln(P_{cross})	-0.047 (0.090)	-0.131 (0.081)	-0.189 (0.073)
R-squared	0.97	0.97	0.96
Shipping dummies	Yes	Yes	Yes
Time \times age dummies	Yes	Yes	Yes
ISBN dummies	Yes	Yes	Yes
n	24,738	24,738	24,738

The dependent variable is the log of the sales rank. This is censored as described in the text. Standard errors are in parentheses. The cross price is the price for the same book at the competitor's site. t_1 denotes April 14 2001. t_2 denotes June 23 2001. t_3 denotes August 3 2001.

data.¹⁵ A second reason to do this is that it gives us a closer look at Amazon.com's pricing changes the summer of 2001. We present these results for three cases—the change in log ranks between periods 1 and 2, the change in log ranks between periods 2 and 3 and the “long-difference” estimator comparing the change in log ranks between periods 1 and 3. The BN.com results use the trimmed LAD.

Table 6 presents the results. Interestingly, the sum of the own price elasticity at each site plus the cross-price elasticity at the other site do approximately equal the same value (as assumed in our specification of equation (3)). However, this conceals an extreme difference in the source of the relative price sensitivity across the two sites. BN.com has a large own price elasticity with a small cross-price from Amazon.com. Amazon.com has the reverse. With the Pareto parameter of 1.2, BN.com's own-price elasticity of demand is around -3.5 . At Amazon.com, on the other hand, it is actually less than one in absolute value, at -0.45 .

The low price elasticity of demand at Amazon.com is important and puzzling. Of course, standard calculations for static imperfectly competitive markets suggest that

15 Trimmed LAD estimation panel procedures for data sets with more than two time periods are not well-developed. The survey of Chay and Powell (2001), for example, present trimmed LAD results only for pairs of time periods rather than for the entire panel.

a firm should choose prices such that the elasticity of demand exceeds 1 in absolute value. However, we are not the first to obtain results estimates of relatively inelastic demand for retail establishments. For example, Hoch et al. (1995) estimate store-level price elasticities of demand of less than 1 in absolute value for many stores in the Dominick's supermarket chain. Note, however, that a firm maximizing dynamic profits might choose a price below this static profit-maximizing level.¹⁶ Prices below the single-period profit-maximizing level would be attractive in a growing market with consumer switching costs, for example. This possibility has been raised in the popular press, where speculation abounds as to whether Amazon.com's prices are sustainable or are artificially low (see, for example, Hansell, 2001). When Amazon.com's growth stops, we may see prices rise substantially.

A second factor to consider is that a 1% increase in the price at Amazon.com reduces quantity by about 0.5% at Amazon.com but raises quantity at BN.com by 3.5%. Given that Amazon.com sells somewhere between three and 10 times as many books as BN.com, this is very close to the same number of books, implying that every customer lost by Amazon.com instead buys the book at BN.com. This is likely to be an unrealistically high degree of switching but the data, at the least, seem to suggest that the cross-price effect is important. The reverse is not true, however. Raising prices by 1% at BN.com reduces sales about 4% but increases sales at Amazon.com by only about 0.2%. Many of the lost customers from BN.com evidently do not just go buy the book from Amazon.com.

Recall that previous research on Internet bookselling has used price dispersion as a proxy for market power. Recalling Table 3, we showed that price dispersion between sellers is quite volatile over the three time periods. Using the price dispersion

Table 7. OLS robustness estimates for BN.com results.

Dep Var.:	(1)	(2)	(3)
ln(Rank)	t_2, t_1	t_3, t_2	t_3, t_1
<i>BN.com (OLS)</i>			
ln(P_{own})	2.399 (0.076)	2.813 (0.074)	2.693 (0.080)
ln(P_{cross})	0.232 (0.045)	-0.222 (0.029)	-0.774 (0.054)
Shipping dummies	Yes	Yes	Yes
Time \times age dummies	Yes	Yes	Yes
ISBN dummies	Yes	Yes	Yes
n	19,126	20,106	19,142

The dependent variable is the log of the sales rank. The data include only those observations in which the BN.com sales ranks are available. Standard errors are in parentheses. The cross price is the price for the same book at the competitor's site. t_1 denotes April 14 2001. t_2 denotes June 23 2001. t_3 denotes August 3 2001.

¹⁶ See Klemperer (1987) and Chevalier and Scharfstein (1996) for discussion.

Table 8. Two period panel estimates of online book demand system by book type.

Dep Var.: ln(Rank)	(1) Amazon.com	(2) BN.com
Time periods t_3, t_1		
ln(P_{own}) Fiction	0.491 (0.129)	3.732 (0.471)
ln(P_{own}) Nonfiction	0.3469 (0.054)	3.142 (0.150)
ln(P_{cross}) Fiction	-0.419 (0.232)	-2.982 (0.489)
ln(P_{cross}) Nonfiction	-0.160 (0.078)	-2.482 (0.145)

The dependent variable is the log of the sales rank. This is censored as described in the text. Standard errors are in parentheses. The cross price is the price for the same book at the competitor's site. t_1 denotes April 14 2001. t_2 denotes June 23 2001. t_3 denotes August 3 2001.

data to infer market power would lead one to have differing conclusions about the degree of market power for each of the three periods. However, when we use our “quantity” data to estimate price elasticities, we observe very little change in the price sensitivity/market power of the two merchants across the different time periods, despite large shifts in the measured dispersion.

We checked the robustness of the trimmed LAD estimator in Table 7. There, we rerun the specifications for BN.com, using OLS in place of the trimmed LAD. We drop observations for which the sales rank is unobserved. These estimates echo the large own price elasticities for BN.com, but show much smaller cross price elasticities of BN.com with Amazon.com.

We also checked the robustness of our results across books of different types and found little difference in the panel specifications of including category-price interaction terms. Table 8 shows the “long difference” panel specifications for both Amazon.com and BN.com, allowing fiction and non-fiction books to have different own- and cross-price elasticities of demand. While fiction books appear slightly more own- and cross-price elastic in both sites' specifications, differences in the coefficients for fiction and non-fiction books are not statistically different from one another. Across our specifications, there was little evidence that the price elasticity of demand varied systematically by type of book. We tried many categories of books as well as formats. All showed the same basic results.

6. Conclusion

This paper has used publicly available data on the prices and sales ranks of more than 18,000 different books at Amazon.com and BN.com to estimate price indices and the amount of price competition online. To do this, we develop a method of

converting sales ranks into actual quantity measures for every book. The results using such data indicate that prices were much more variable online than in retail stores during the time period of this sample and point to an important outlet substitution bias in the CPI for recreational books over this time period. Second, the results show that there is significant price sensitivity of online customers both to a site's own price as well as to leading rival's price. This is much more true at BN.com, however, where the own price elasticity of demand is close to -4 and the cross-price elasticity very high, than at Amazon.com where the price elasticity is around -0.6 , and the cross-price elasticity is relatively small. The results also show that using price dispersion as a proxy for market power is not appropriate in our data.

Taken together, our results point to Amazon.com as a clear market leader in the online book business with BN.com serving as more of a price-taking fringe. The usefulness of the sales rank data in allowing us to actually estimate the degree of market power in markets with little publicly available quantity data raises the question of whether similar information could be gathered for other industries.

Acknowledgments

We would like to thank Chris Forman, Michael Smith, Scott Schaefer, Joel Waldfogel and numerous seminar participants for helpful comments. We benefited greatly from expert research assistance by Chip Hunter, Patrik Guggenberger, and Tina Lam. We thank Madeline Schnapp for assistance and rely heavily on her 6/21/2001 presentation at the UCB/SIMS Web Mining workshop for data. Goolsbee would like to thank the National Science Foundation (SES 9984567), the Sloan Foundation and the American Bar Foundation for financial support. We have no reason to thank the University of Chicago Press.

References

- Amazon.com, e-mail correspondence with customer service agent John Armstrong, May 15 2000.
- American Booksellers Association, Industry News, "Overall Book Sales Up Slightly for First Six Months of '01", November 1 2001, <<http://www.bookweb.org/home/news/btw/5182.html>>, accessed 5/23/02.
- Bailey, Joseph P. (1998). "Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce", Ph.D. dissertation, Department of Electrical Engineering and Computer Science, MIT.
- BN.com, e-mail correspondence with customer service agent Charlie, January 14 2000.
- Boskin, Michael J., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson. (1996). "Toward a more accurate measure of the cost of living", Final Report to the Senate Finance Committee.
- Boston Consulting Group, The State of Online Retailing 3.0, with Shop.org November 2000.
- Brown, Jeffrey and Austan Goolsbee. (2002). "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry", *Journal of Political Economy* 110(3), 481-507.
- Brynjolfsson, Erik and Smith, Michael. (2000). "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", *Management Science* 46, 563-585.
- Cader, Michael. (2001). "Online Bookselling: The New Mythology in the Making", Publisher's Lunch e-mail newsletter April 16 2001.

- Cage, Richard. (1996). "New Methodology for Selecting CPI Outlet Samples", *Monthly Labor Review*, December 1996, <<http://www.bls.gov/cpi/cpirc001.htm>>, accessed 7/3/2002.
- Carlton, D. and J. Chevalier. (2001). "Free Riding and Sales Strategies for the Internet", *The Journal of Industrial Economics* 49, 441-462.
- Chay, Kenneth and James L. Powell. (2001). "Semiparametric censored regression models", *Journal of Economic Perspectives* 15, 29-42.
- Chevalier, Judith A. and David Scharfstein. (1996). "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence", *American Economic Review* 86, 703-725.
- Clay, Karen, Ramayya Krishnan, and Eric Wolff. (2001). "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry", *The Journal of Industrial Economics* 49, 521-540.
- Ellison, Glenn and Sara Fisher Ellison. (2001). "Search, Obfuscation, and Price Elasticities on the Internet", MIT working paper.
- Goolsbee, Austan. (1999). "Evidence on the High Income Laffer Curve from Six Decades of Tax Reform", *Brookings Papers on Economic Activity* 1999(2), 1-47.
- Goolsbee, Austan. (2000). "In a World Without Borders: The Impact of Taxes on Internet Commerce", *Quarterly Journal of Economics* 115, 561-576.
- Goolsbee, Austan. (2001). "Competition in the Computer Industry: Online vs. Retail", *Journal of Industrial Economics* 49, 487-499.
- Hansell, Saul. (2001). "Listen Up! It's Time for a Profit, A Front-Row Seat as Amazon Gets Serious", *New York Times*, May 20 (2001).
- Hoch, Stephen J., Kim, Byung-Do, Montgomery, Alan, and Rossi, Peter. (1995). "Determinants of Store-level Price Elasticity", *Journal of Marketing Research* 32, 17-29.
- Honore, Bo. (1992). "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models", *Econometrica* 60, 533-565.
- Italie, Hilltel. (2001). "Amazon's Bottom 10: Not Exactly Page Turners" *Chicago Sun-Times*, August 17 2001.
- Johnson, Norman L. and Samuel Kotz. (1970). *Distributions in Statistics: Continuous Univariate Distributions*. Vol. 1, Houghton-Mifflin, Boston.
- Klemperer, Paul. (1987). "Markets with consumer switching costs", *Quarterly Journal of Economics* 102, 375-394.
- Kosala, Raymond and Hendrik Blockeel. (2000). "Web Mining Research: A Survey", *SIGKDD Explorations* 2, 1-15.
- Kuttner, Robert. (1998). "The Net: A Market Too Perfect for Profits", *Business Week* 20, May 11 1998.
- Lee, Ho Geun. (1997). "Do Electronic Marketplaces Lower the Price of Goods?" *Communications of the ACM* 41, 73-80.
- Mandelbrot, Benoit B. (1983). *Fractal Geometry of Nature*. W.H. Freeman and Company.
- Pareto, V. (1897). *Cours d'Economie Politique*. F. Rouge, Lausanne.
- Poynter, Daniel. (2002). "Publishing Poynters", April-June 2000, <<http://parapub.com/getpage.cfm?file=newsletter/News0400.html&userid=1035600>>, accessed June 5 2002.
- Quandt, R.E. (1964). "Statistical discrimination among alternative hypotheses and some economic regularities", *Journal of Regional Science* 5, 1-23.
- Rayport, Jeffrey F., Knoop, Carin-Isabel, and Reavis, Cate. (1998). "Selling Books Online in Mid-1998", HBS Case 9-899-038.
- Reinsdorf, Marshall. (1993). "Price Dispersion, Seller Substitution, and the U.S. CPI", *Bureau of Labor Statistics Working Paper* 252.
- Schnapp, Madeline and Tim Allwine. (2001). "Mining of book data from Amazon.com", Presentation at the UCB/SIMS web mining conference, <http://www.sims.berkeley.edu/resources/affiliates/workshops/webmining/slides/ORA.ppt>.
- Schultze, Charles and Christopher Mackie (ed.) (2002). *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Committee on National Statistics, National Research Council, Washington, D.C.: National Academy Press.

- Scott Morton, Fiona, Florian Zettlemeyer, and Jorge Silva-Russo. (2001). "Internet Car Retailing", *Journal of Industrial Economics* 49, December 2001, 501–519.
- Smith, Michael and Erik Brynjolfsson. (2001). "Consumer Decision-making at an Internet Shopbot: Brand Still Matters", *The Journal of Industrial Economics* 49, December 2001, 541–558.
- Weingarten, Gene. (2001). "Below the Beltway", *Washington Post* June 17 2001.
- White, E. (2000). "No Comparison. Shopping Bots were Supposed to Unleash Brutal Price Wars. Why Haven't They?" *Wall Street Journal*, October 23 2000, R18.
- Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, Cambridge.