



---

The Effect of Word of Mouth on Sales: Online Book Reviews

Author(s): Judith A. Chevalier and Dina Mayzlin

Source: *Journal of Marketing Research*, Vol. 43, No. 3 (Aug., 2006), pp. 345-354

Published by: [American Marketing Association](#)

Stable URL: <http://www.jstor.org/stable/30162409>

Accessed: 06/03/2014 16:57

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Marketing Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Marketing Research*.

<http://www.jstor.org>

JUDITH A. CHEVALIER and DINA MAYZLIN\*

The authors examine the effect of consumer reviews on relative sales of books at Amazon.com and Barnesandnoble.com. The authors find that (1) reviews are overwhelmingly positive at both sites, but there are more reviews and longer reviews at Amazon.com; (2) an improvement in a book's reviews leads to an increase in relative sales at that site; (3) for most samples in the study, the impact of one-star reviews is greater than the impact of five-star reviews; and (4) evidence from review-length data suggests that customers read review text rather than relying only on summary statistics.

## The Effect of Word of Mouth on Sales: Online Book Reviews

Online user reviews have become an important source of information to consumers, substituting and complementing other forms of business-to-consumer and offline word-of-mouth communication about product quality. Consequently, many managers believe that a Web site must provide community content to build brand loyalty (see, e.g., Fingar, Kumar, and Sharma 2000; McWilliam 2000). Despite this widespread belief, to our knowledge, there is no literature documenting that community content plays any role in consumer decision making. It seems that such a finding is a necessary prerequisite for content provision to be a profitable strategy.

There are many reasons to suspect *ex ante* that creating a forum for community content could be a poor strategy. First, it is not clear why users would bother to take the time to provide reviews for which they are not in any way compensated.<sup>1</sup> Second, competing retailers can free ride on investments in recommender systems; there is nothing to stop a consumer from using the information provided by one Web site to inform purchases made elsewhere. Third, by providing user reviews, a site cedes control over the

information displayed; unfavorable reviews may depress sales. This may be less of a threat to a retailer that sells many different brands than to a manufacturer. Similarly, because interested parties can freely proliferate favorable reviews of their own products, positive reviews may not be credible and may not function to stimulate sales.<sup>2</sup> Finally, online user reviews may not be useful and may not stimulate sales because of the sample selection bias that is inherent in an amateur review process. That is, a consumer chooses to read a book or watch a movie only if he or she believes that there is a high probability of enjoying the experience. In the presence of consumer heterogeneity, this implies that the pool of reviewers will have a positive bias in their evaluation compared with the general population. Thus, positive reviews may simply be discounted by potential buyers.<sup>3</sup>

In this study, we characterize patterns of reviewer behavior and examine the effect of consumer reviews on firms' sales patterns. In particular, we use publicly available data from the two leading online booksellers, Amazon.com and Barnesandnoble.com (bn.com), to construct measures of each firm's sales of individual books. Both bn.com and Amazon.com allow customers to post reviews on the site. Our econometric analysis is designed to answer the following question: If a cranky consumer posts a negative review of a book on bn.com but not on Amazon.com, would the sales of that book at bn.com fall relative to the sales of that book at Amazon.com? To isolate the answer to this question, we propose a "differences-in-differences" approach. For a sample of books, we measure reviews and a proxy for sales at Amazon.com and bn.com over three time points. We examine whether a change in the number and valence of

<sup>1</sup>Steven Levitt ponders this question at length in the Freakonomics blog (<http://www.freakonomics.com/2005/07/why-do-people-post-reviews-on-amazon.html>).

\*Judith A. Chevalier is William S. Beinecke Professor of Finance and Economics (e-mail: [judith.chevalier@yale.edu](mailto:judith.chevalier@yale.edu)), and Dina Mayzlin is Assistant Professor of Marketing (e-mail: [dina.mayzlin@yale.edu](mailto:dina.mayzlin@yale.edu)), Yale School of Management. The authors thank the two anonymous *JMR* reviewers for comments that greatly improved the article. They also thank participants at many seminars for helpful comments, as well as Sharon Oster, Jackie Luan, David Godes, and Jiwoong Shin. The authors are especially grateful for the comments and encouragement of the late Dick Wittink. Jessie Cheng and Tudor Olteanu provided excellent research assistance. Both authors contributed equally, and their names are listed in alphabetical order. Pradeep Chintagunta served as editor for this article.

<sup>2</sup>For a theoretical treatment of recommendation systems on which firms can anonymously post reviews, see Mayzlin (2006).

<sup>3</sup>In a different context, Resnick and Zeckhauser (2002) find that 99% of the feedback ratings on eBay.com are positive.

reviews over time for a particular book at one site relative to the other site predicts a change in the subsequent sales of that book at one site relative to the other. By focusing on the differences between the relative sales of the book at the two sites, we are able to control for the possible effect of unobserved book characteristics on both reviews and sales. By focusing on the differences across sites over time, we control for the possibility that taste differences across the customer populations at the two sites differ in a way that affects both reviews and sales.

Our findings suggest that, on average, reviews tend to be positive, especially at bn.com. We show that the addition of new, favorable reviews at one site results in an increase in the sales of a book at that site relative to the other site. We find some evidence that an incremental negative review is more powerful in decreasing book sales than an incremental positive review is in increasing sales. Our results on the length of reviews suggest that consumers actually read and respond to written reviews, not merely the average star ranking summary statistic provided by the Web sites.

We organize the rest of the article as follows: First, we describe the data. Second, we describe the methodology. Third, we present the results of the cross-sectional and differences-in-differences analyses of the effect of reviews on sales. Finally, we conclude.

#### DATA

Our data consist of individual book characteristics and user review data collected from the public Web sites of Amazon.com and bn.com. Our goal was to generate a representative sample of sales. Because we did not have access to the firms' proprietary sales data, we approximated a random sample of sales as follows: First, we collected a random sample of 3587 books from *Global Books in Print* (see [www.GlobalBooksInPrint.com](http://www.GlobalBooksInPrint.com)) that were released over the 1998–2002 period. However, titles chosen at random are likely to have low sales because a large fraction of sales are concentrated in a small fraction of books. It is possible that word of mouth may be especially influential on the sales of these books because there are few other sources of information on these titles. Thus, in addition, we collected data on all 2818 titles that appeared in *Publishers Weekly* best-seller lists from January 14, 1991, to November 11, 2002 (see [www.publishersweekly.com](http://www.publishersweekly.com)), a period ending approximately six months before our data collection.

We collected data during three periods: for a two-day period in May 2003, for a two-day period in August 2003, and for a two-day period in May 2004. For each book in our sample at each time, we gathered the price charged for the book, the promised time until the book would ship, the number of reviews, and the average number of stars the reviewers assigned (on a scale of one to five stars, with five stars being the best). Most of the books have a promised delivery of 24 hours (96% at Amazon.com and 88% at bn.com). However, Amazon.com and bn.com use other shipping categories, such as "Usually ships in 2–3 days" or "Special order: usually ships in 1–2 weeks."

For all periods, we extracted detailed characteristics of the most recent 500 reviews of the book posted on the Web site, including the number of stars assigned and the date the review was posted.<sup>4</sup> We also extracted the "sales rank" of

each book at each site. At each site, the top-selling book at that site has a sales rank of one, and the lower sellers are assigned higher sequential ranks. We included in our data only books listed as "available" at Amazon.com and bn.com. Not surprisingly, many of the books drawn randomly from the *Global Books in Print* sample were not available for sale on the Web sites. At the first period, we found these basic data for 1909 of the *Global Books in Print* sample of books and for 2261 of the past decade's best-sellers sample.

For each book in our sample, we identified all formats of that book (audio, paperback, hardback, large print, and so forth). We excluded audio books. Amazon.com and bn.com provide identical reviews for all the different formats of a given title. In general, there is one format that is extremely dominant. Because we did not want the data set to include duplicate information, we examined sales and reviews only for the most popular format within a title. We then excluded from our analysis books for which the most popular format within the title was different at Amazon.com and bn.com. For example, if the hardcover was the better seller at Amazon.com and the paperback was the better seller at bn.com, we excluded the book from our sample.<sup>5</sup>

Chevalier and Goolsbee (2003a) report that Amazon.com claims that for books in the top 10,000 ranks, the rankings are based on the last 24 hours and are updated hourly. For books ranked 10,001–100,000, the ranks are updated once a day. For books ranked greater than 100,000, the sales ranks are updated once a month (Amazon.com 2000). Based on this system, books that have not been purchased in the past month would not be ranked. However, many hundreds of thousands of books have a rank but almost certainly have fewer than one sale per month. Italic (2001) claims that for these rarely purchased books, Amazon.com bases the rank on the total sales since Amazon's inception. Barnesandnoble.com claims to update all of its rankings daily (bn.com 2000).<sup>6</sup> Thus, with the exception of the books that have very high ranks (low sales) on Amazon.com, the rankings represent a current snapshot of sales. However, bn.com provides only sales ranks for approximately 650,000 books. There are books at bn.com that are available for purchase but for which the rank is "too high" (sales are too low) to be disclosed. Amazon.com does not censor its sales ranks, and they appear to range upward of one million. If we were to use as our sample all books with prices and ranks at both sites, our sample would contain a large number of books that are relatively popular at bn.com and relatively unpopular at Amazon.com. However, books that are relatively popular at Amazon.com and relatively unpopular at bn.com would not appear in the sample, because they have been censored out by bn.com's rank-reporting strategy. To address this asymmetry, we removed from our sample books with ranks greater than 650,000 at Amazon.com. More important, removing these books serves to remove books for which the Amazon.com ranks are updated infrequently. The final sample contained 2387 observations,

<sup>5</sup>The main results are qualitatively robust to several format selection criteria.

<sup>6</sup>Because 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. Despite repeated requests, bn.com would not provide us with any more detail on its ranking system.

<sup>4</sup>For the first period only, we also extracted the full text of the most recent 500 reviews posted on the Web site.

1087 of which had reviews posted at both sites at the first (May 2003) time point.

We examined differences in sales over the May 2003–August 2003 horizon and over the May 2003–May 2004 horizon. As we explained previously, sales ranks at a particular moment in time represent a snapshot of sales for up to a month. Thus, we decided to use a conservative approach to measure the rank–sales relationship. We also examined the relationship between changes in sales over the May 2003–August 2003 period and changes in reviews over the May 2003–July 2003 period. We examined the relationship between changes in sales over the May 2003–May 2004 period and changes in reviews over the May 2003–April 2004 period. Because reviews are dated, we could extract the appropriate sample of reviews from the data collected at the August 2003 and May 2004 periods.

Thus, we rely on sales-ranking data rather than more conventional sales data. For most of our analysis, we simply use the sales ranks directly in our analysis and discuss the impact of reviews on sales ranks rather than sales. However, Schnapp and Allwine (2001), Chevalier and Goolsbee (2003a), and Rosenthal (2005) all find that the relationship between  $\ln(\text{sales})$  and  $\ln(\text{ranks})$  is approximately linear. Using Schnapp and Allwine's methodology enables us to translate sales ranks into sales approximations and thus calibrate the relationship between reviews and sales.<sup>7</sup>

<sup>7</sup>This previous literature approximates that for Amazon.com,  $\ln(\text{sales}) = 9.61 - .78\ln(\text{rank})$ . For bn.com, in line with the work of Chevalier and Goolsbee (2003b), we scale the relationship down to capture the fact that its sales are 15% of Amazon.com. In addition, we control for the 24%

Table 1 presents the summary statistics for the main variables of interest in our data. The number of observations shrinks across time because books must be available at Amazon.com and bn.com in the first period to be included in the first period's sample, but they must be available at both sites in both the first period and the subsequent period for each of the other two samples. The most striking finding in Table 1 is how positive the reviews are at both sites and at all times. For all of our time points and at both sites, the modal review is five stars, and the mean number of stars for any book (with reviews) is greater than four stars.

However, there are a few notable differences across the sites that are apparent in Table 1. We highlight three: (1) For the books in our sample, bn.com prices are significantly higher (as can be shown in a paired t-test); (2) Amazon.com has more reviews than bn.com, and bn.com has a much higher fraction of books in our sample that have no reviews at all (54% for bn.com versus 13% for Amazon); and (3) reviews are slightly more positive on average at bn.com, though again, they are overwhelmingly positive overall at both sites.<sup>8</sup>

Across time, we do not note big changes in pricing or reviewing behavior. Book rankings increase as book popularity declines. This leads to books dropping out of the sam-

growth in Amazon.com in the intervening years. The final relationships are  $9.825 - .78\ln(\text{rank})$  for Amazon.com and  $\ln(\text{sales}) = 7.928\ln(\text{rank})$  for bn.com.

<sup>8</sup>Although bn.com is currently more expensive than Amazon.com, this has not always been the case historically (see, e.g., Chevalier and Goolsbee 2003a).

Table 1  
SUMMARY DATA

	May 2003		August 2003		May 2004	
	Amazon.com	bn.com	Amazon.com	bn.com	Amazon.com	bn.com
Price	13.97 (14.41)	15.50 (14.75)	13.85 (14.84)	15.2 (15.28)	13.56 (15.12)	15.22 (15.79)
Ranking	129,799 (169,363)	121,061 (156,903)	134,303 (166,575)	122,377 (152,466)	123,112 (152,349)	137,402 (166,939)
Number of reviews per book	60.99 (180.40)	12.79 (44.55)	59.79 <sup>a</sup> (183.70)	13.11 (46.70)	68.31 (205.42)	14.15 (42.30)
Average stars	4.14 (.70)	4.45 (.57)	4.13 (.71)	4.16 (.62)	4.06 (.70)	4.43 (.58)
Fraction of one-star reviews	.07 (.12)	.03 (.08)	.07 (.12)	.03 (.08)	.08 (.12)	.04 (.09)
Fraction of five-star reviews	.57 (.29)	.67 (.26)	.57 (.29)	.67 (.26)	.51 (.29)	.66 (.26)
Incremental reviews per book			1.82 (5.49)	.53 (2.27)	10.56 (67.59)	1.85 (18.72)
Change in average stars			-.010 (.16)	-.010 (.14)	-.006 (.50)	-.015 (.27)
Fraction of books with no reviews	.13	.54	.12	.54	.17 <sup>b</sup>	.49
Number of observations	2387	2387	2082	2082	1636	1636

<sup>a</sup>This number is slightly lower than the average number of reviews per book in the May 2003 Amazon.com sample. This is not due to a loss of reviews over this period but rather to the change in the sample. A few of the books that had a high number of reviews did not have rank information in August 2003; thus, we did not include them in the sample.

<sup>b</sup>The fraction of books with no reviews on Amazon.com goes up in this period. This is at least partially due to the pruning of reviews by Amazon.com, which we discuss in further detail herein.

Notes: The sample comprises all books in our database with complete data and with an Amazon.com rank of less than 650,000, for which the most popular format of the book at Amazon.com is the same as the most popular format of the book at bn.com. Means are primary data entry, and standard deviations are in parentheses.

ple, and thus the summary statistic rankings do not change much over time. The number of incremental reviews posted for each book between May 2003 and July 2003 (recall that we measure sales changes from May to August, but we measure review changes from May to July) is small. In the first two months, the average book in our sample picks up an additional review at Amazon.com and an additional half of a review at bn.com. However, over the longer horizon, more reviews are posted. The typical book gains 11 reviews at Amazon.com and 2 at bn.com over the 11-month reviewing horizon. The data do not suggest that prior reviews for a given book are systematically more or less enthusiastic than subsequent ones. The mean change in average star rating of a book between May 2003 and April 2004 is within one standard deviation of zero.

In addition to the actual ratings the reviewers give, there might be additional information contained in the message text. Unfortunately, reading the reviews is an extremely costly task, and the measures obtained are very noisy, as Godes and Mayzlin (2004) show. Text analysis programs are imperfect.<sup>9</sup> However, a relatively cost-effective measure of the review text is the length (total number of typed characters) contained in the review. A priori, it is not completely clear how to interpret this measure. One possibility is that a longer review represents more effort on the part of the reviewer. Another possibility is that a longer explanation is required to support a "mixed" review. We find partial support for the latter interpretation: Table 2 shows the frequency distribution for all types of reviews for the May 2003 sample and shows that for both sites, one-star and five-star reviews are much shorter than two-star, three-star, and four-star reviews. Another pattern that emerges is that Amazon.com reviewers post longer reviews at all star levels than do their peers at bn.com.

#### MODEL SPECIFICATION

Consider book  $i$  that is sold on Amazon.com and bn.com. Ideally, our dependent variable would be log of sales of a book on a particular site. The reason for the log specification rather than levels is that the log specification estimates the effect of a change in the independent variables on the percentage change in the dependent variable. This is appro-

<sup>9</sup>Indeed, it has been suggested that we use these data to train text analysis programs. The idea is that a five-star review must be more enthusiastic than a four-star or three-star review, and the program can use the reviews to glean patterns that measure levels of enthusiasm.

priate because in our case, there are scale effects. Exogenously, a large number of people view the "popular" book's page at Amazon.com and bn.com, and a small number of people view the "unpopular" book's page at Amazon.com and bn.com. The fraction of these viewers who go on to buy is plausibly a function of the reviews posted on the site. Although log sales would be the ideal dependent variable, we use log rank. Moreover, Schnapp and Allwine (2001) use proprietary data on the sales of a sample of books on Amazon.com to map the relationship between ranks and sales; they find that the relationship between log ranks and log sales is close to linear. This finding suggests that in lieu of sales data, log rank is the appropriate dependent variable. Because of the linear relationship between log ranks and log sales, if we were to use our estimate of log sales as the dependent variable, the estimated coefficients in our specifications and their standard errors would simply be scaled by a constant.

The book's sales rank on a site is a function of a book fixed effect ( $v_i$ ), a book-site fixed effect ( $\mu_i^A$ ), and other factors. The book fixed effect is related to factors such as the offline promotion, the quality of the book, and the popularity of the author. The book-site fixed effect is related to the fit between the book and the preferences of the customers of the site. That is,

$$(1) \quad \ln(\text{rank}_i^A) = \mu_i^A + v_i + \alpha^A \ln(P_i^A) + \gamma^A \ln(P_i^B) + X\Gamma^A + S\Pi^A + \varepsilon_i^A, \text{ and}$$

$$(2) \quad \ln(\text{rank}_i^B) = \mu_i^B + v_i + \alpha^B \ln(P_i^B) + \gamma^B \ln(P_i^A) + X\Gamma^B + S\Pi^B + \varepsilon_i^B,$$

where rank denotes the sales rank; the superscripts A and B refer to Amazon.com and bn.com, respectively; P denotes price;<sup>10</sup> X denotes the vector of review variables from both sites (we allow Amazon.com reviews to affect bn.com's customers and bn.com reviews to affect Amazon's customers); and S is a vector of dummy variables summarizing the shipping times promised by each Web site for each book. For each of bn.com and Amazon.com, we have a dummy variable that indicates "usually ships in 24 hours" (the most frequent category), a dummy that indicates "usually ships in 2-3 days," and so forth. For each book, S has a

<sup>10</sup>We take the log of price to estimate the effect of percentage change in price on percentage change in rank.

Table 2  
REVIEW LENGTH AND STAR DISTRIBUTION FOR THE MAY 2003 SAMPLE

	Amazon.com		bn.com	
	Frequency	Number of Typed Characters	Frequency	Number of Typed Characters
One-star reviews	8.97	765	3.44	558
Two-star reviews	7.53	916	4.07	599
Three-star reviews	10.56	997	6.00	566
Four-star reviews	19.89	949	19.27	577
Five-star reviews	53.05	812	67.22	508
Overall		854		529

Notes: The sample includes all books with reviews in May 2003.



1 for the promised ship time category at both Amazon.com and bn.com. We use four possible shipping time categories at Amazon.com and three at bn.com.

Because we expect the unobservable fixed effects to be correlated with independent variables, omitting these effects would bias the coefficients on the review variables.<sup>11</sup> If we assume that the two sites are virtually identical in terms of their readership's preferences (i.e., if  $\mu_i^A = \mu_i^B$ ),<sup>12</sup> we can eliminate the fixed effects by differencing the data across sites:

$$(3) \quad \ln(\text{rank}_i^A) - \ln(\text{rank}_i^B) = \beta^A \ln(P_i^A) + \beta^B \ln(P_i^B) + X\Gamma + S\Pi + \varepsilon_i.$$

However, if there are subtle differences across the two sites (i.e., if  $\mu_i^A \neq \mu_i^B$ ), we need to obtain another data point and difference the data across the sites and across time:

$$(4) \quad \Delta[\ln(\text{rank}_i^A) - \ln(\text{rank}_i^B)] = \beta^A \Delta \ln(P_i^A) + \beta^B \Delta \ln(P_i^B) + \Delta X\Gamma + \Delta S\Pi + \varepsilon_i.$$

The advantage of Equation 3 is that it allows us to use more data because many books' reviews do not change over time. In addition, it allows us to estimate the price coefficients because there is not a great amount of variation in prices across time.<sup>13</sup> However, although the differences-in-differences specification in Equation 4 leaves us with a smaller sample and does not allow us to estimate all the coefficients of interest, it has the advantage of eliminating the book-site-specific fixed effects. If, for example, bn.com users simply like computer books less than Amazon.com users (buying them less and giving them worse reviews), differencing the data would eliminate the problem. Thus, although we briefly present the cross-sectional results, our main focus is Equation 4.

### THE EFFECT OF REVIEWS ON SALES

#### Cross-Sectional Analysis

In this subsection, we assume that there are no site-specific fixed effects and examine the relationship between a book's customer reviews and its sales rank across sites (see Equation 3). Table 3 presents the estimation results for this sample. Table 3, Column 1, presents the results for a regression in which no review variables are included, only prices at both sites and the shipping dummies. The price coefficients reflect a combination of own- and cross-price elasticities at both sites. The price coefficient for Amazon.com is positive and statistically significant, suggesting that when prices rise, sales ranks at Amazon.com become larger (i.e., sales fall). The price coefficient is negative for bn.com. This is as expected; recall that the left-

hand-side variable is  $\ln(\text{rank})$  at Amazon.com minus  $\ln(\text{rank})$  at bn.com. Again, when prices rise at bn.com, sales ranks become larger (i.e., sales fall at bn.com relative to Amazon.com). The absolute value of the price coefficient is larger at bn.com, suggesting that sales ranks respond more to prices at bn.com than at Amazon.com. This is consistent with Chevalier and Goolsbee's (2003a) findings that demand is more elastic at bn.com than at Amazon.com.

Table 3, Column 2, includes measures of the total number of reviews for each book and the average star ranking of each book's reviews. Specifically, we include the natural log of the total number of reviews at Amazon.com and the natural log of the total number of reviews at bn.com. These are set to zero when the number of reviews equals zero. We also include two dummies: one that takes the value of 1 when a title at Amazon.com has no reviews (and 0 otherwise) and one that takes the value of 1 when bn.com has no reviews (and 0 otherwise). Finally, we include the average star value of the book's customer reviews at each site in the regression.

As we expected, for both sites, the coefficients for the average star value suggest that sales improve when books are rated more highly, but the effect is statistically insignificant for bn.com. To illustrate the magnitude of the effects, we consider a book with four five-star reviews at both Amazon.com and bn.com and a rank of 500 at both sites. Imagine that one of the five-star reviews at Amazon.com was changed to a one-star review. Given the relationship between ranks and sales, the coefficients imply that if bn.com's ranking of the book were unchanged by this review change, the rank at Amazon.com would be expected to rise to 601, an estimated decrease in sales of approximately 20 books per week. Another useful way to interpret the coefficient magnitude is to consider the impact of a review on a book that has no reviews on either sites. Our estimates imply that if the book receives one Amazon.com review with one, two, or three stars, its rank on Amazon.com will rise (sales fall), assuming that its rank on bn.com stays constant. However, if the book receives a positive review of four or five stars, its rank on Amazon.com will fall (sales rise).

Table 3, Column 3, focuses on a different way of measuring review valence. In place of average stars, the fraction of reviews that are one-star reviews and the fraction of reviews that are five-star reviews are included for each site. As we expected, the coefficients suggest that five-star reviews improve sales and one-star reviews hurt sales in a statistically significant way at Amazon.com. The coefficient for one-star reviews for bn.com is of the expected sign and statistically significant at the 7% level. However, the coefficient for five-star reviews is almost zero but of the "wrong" sign. Nonetheless, the one-star reviews have large coefficients in absolute value relative to the five-star reviews, indicating that the relatively rare one-star reviews carry a lot of weight with consumers. This result also makes sense when the credibility of one-star and five-star reviews is considered. After all, the author, or another interested party, may "hype" his or her own book by publishing glowing reviews on these Web sites.<sup>14</sup> Although the author can post

<sup>11</sup>In addition, note that the correlation between review variables and the fixed effect induces dependence in review variables over time. However, although this implies that the right-hand-side variables may be correlated in the differences-in-differences specification, it does not bias the estimation results.

<sup>12</sup>We have some evidence that the two sites' readers and reviewers exhibit similar preferences. For example, we find that the correlation between log ranks of individual books is high (.825 for the 2387 books in our first sample). We also do not find differences in review patterns across sites that are subject specific. For example, juvenile fiction received the highest reviews and serious nonfiction received the lowest reviews on both sites (for more details, see Chevalier and Mayzlin 2003).

<sup>13</sup>This also allows us to compare our results with previous work.

<sup>14</sup>For one well-publicized, alleged example in economics, see Morin (2003).

Table 3  
THE EFFECT OF REVIEWS ON SALES

	1	2	3	4	5
Amazon.com ln(price)	1.556*** (.159)	1.545*** (.155)	1.532*** (.156)	2.147*** (.324)	2.148 (.328)
bn.com ln(price)	-1.801*** (.148)	-1.837*** (.144)	-1.826*** (.145)	-2.67*** (.280)	-2.58 (.282)
Amazon.com ln(number of reviews)		-.215*** (.024)	-.205*** (.023)	-.403*** (.050)	-.373 (.050)
bn.com ln(number of reviews)		.131*** (.033)	.13*** (.033)	.259*** (.052)	.242 (.052)
Amazon.com no-reviews dummy		-.574*** (.187)	.075*** (.109)		
bn.com no-reviews dummy		-.154 (.100)	-.354** (.131)		
Amazon.com average star rating		-.184*** (.038)		-.418*** (.079)	
bn.com average star rating		.024 (.017)		.145* (.088)	
Amazon.com fraction of five-star reviews			-.256*** (.100)		-.704*** (.235)
bn.com fraction of five-star reviews			-.147 (.149)		.061 (.188)
Amazon.com fraction one-star reviews			.483** (.255)		1.15** (.506)
bn.com fraction of one-star reviews			-.836* (.467)		-.94* (.566)
Number of observations	2387	2387	2387	1087	1087
Includes shipping dummies?	Yes	Yes	Yes	Yes	Yes
R-square	.086	.138	.136	.216	.203

\* $p < .10$ .

\*\* $p < .05$ .

\*\*\* $p < .01$ .

Notes: In Columns 1–3, the sample is the complete May 2003 sample. In Columns 4 and 5, the sample is the books that had at least one review on both sites in May 2003. The dependent variable is the difference between the log ranking of the book on Amazon.com and the log sales ranking of the book on bn.com. That is, the dependent variable is  $\text{Ln}(\text{rank}^A) - \text{Ln}(\text{rank}^B)$ .

a large number of meaningless five-star reviews cheaply, he or she cannot prevent others from posting one-star reviews.<sup>15</sup>

We examine the robustness of these estimates in Table 3, Columns 4 and 5. In particular, in Column 4, we repeat the specification of Column 2, but we examine only the subsample of 1087 books that have at least one review on each site. We drop the “no-review” variables but measure the impact of number of reviews and star rankings for this subsample. The results are similar to those we presented previously. All the signs of the coefficients of interest are as we predicted. The coefficient magnitudes and significance levels for the variables measuring star ratings are somewhat larger than in the full sample.

Finally, we use the cross-sectional sample to examine the relationship between review lengths and sales. To do this, we repeat the specifications in Table 3, Columns 4 and 5, including the natural log of the average length of all the

reviews for each book at each site. The results appear in Table 4. The coefficient for review length is positive and statistically significant at Amazon.com, and it is negative and insignificant at bn.com. This suggests that when we control for the star rating of the book, longer reviews depress the site’s relative share.

There are (at least) two possible interpretations of this result. The first, which we view as the less likely, is that encouraging longer, more useful, and more nuanced reviews is actually harmful to sales. However, it is more likely that within each site, the length of the review is correlated with the enthusiasm of the review in ways that are not captured by the star measures. For example, even within the realm of the statistically dominant 5-star reviews, there could be differing degrees of enthusiasm. That is, some “read like” 4.5-star reviews, while some read more like 5-star reviews. Those that read like 4.5-star reviews might be longer on average because they are more likely to be mixed (i.e., both negative and positive aspects of the book are discussed). We find some evidence for this in our data. Consider the subsample of 1087 books with at least one review at both sites. Within that group, consider the subsample of 5-star reviews. The average length of 5-star reviews at Ama-

<sup>15</sup>It could be argued that posting one-star reviews for competing books could be a reasonable strategy for an author. We acknowledge that this may be true, though it is not at all clear that two books on the same subject, for example, are substitutes rather than complements.

Table 4  
THE EFFECT OF REVIEW LENGTH ON BOOK MARKET  
SHARES

	1	2
Amazon.com ln(price)	2.127** (.325)	2.093** (.326)
bn.com ln(price)	-2.661** (.279)	-2.634** (.280)
Amazon.com ln(number of reviews)	-.415** (.0501)	-.411** (.0503)
bn.com ln(number of reviews)	.267** (.0518)	.267** (.052)
Amazon.com average star rating	-.405** (.0794)	
bn.com average star rating	.138* (.0878)	
Amazon.com fraction of five-star reviews		-.441* (.242)
bn.com fraction of five-star reviews		.083 (.188)
Amazon.com fraction of one-star reviews		1.550** (.511)
bn.com fraction of one-star reviews		-1.020* (.563)
Amazon.com ln(average review length)	.570** (.146)	.598** (.151)
bn.com ln(average review length)	-.049 (.0917)	-.052 (.0920)
Number of observations	1087	1087
Includes shipping dummies?	Yes	Yes
R-square	.217	.216

\* $p < .10$ .

\*\* $p < .01$ .

Notes: The sample is the books that had at least one review on both sites in May 2003. The dependent variable is  $\ln(\text{rank}^A) - \ln(\text{rank}^B)$ .

zon.com is 795 characters for books with the average star rating of 4 or greater, and it is 847 characters for books with the average star rating of less than 4. Similarly, the average review length at bn.com is 492 for 5-star reviews for a book with the average rating of 4 or greater, and it is 675 for 5-star reviews for a book with the average rating of less than 4. If we assume that the books with the lower average ratings have the "less enthusiastic" 5-star reviews, this at least suggests that even within the 5-star category, review length is correlated with the reviewer's level of enthusiasm for the book. Regardless of the interpretation of the length results, the results seem to suggest that customers read and respond to the review content at each site. However, longer reviews do not necessarily stimulate sales.

#### Differences-in-Differences Analysis

As we discussed previously, omitted book-site fixed effects could bias the preceding results. To eliminate a possible site-specific fixed effect, we collected review data for May 8 and July 8, 2003, and the ranks, prices, and shipping data as posted on August 8, 2003, for the sample of 2387

books analyzed previously.<sup>16</sup> The specification we estimate is given in Equation 4.

Of the sample of 2387 books, only 2082 books were available at both sites in the second period and contained rank information at both sites. This short differences-in-differences time window is useful because it is likely that the underlying characteristics of site users remain relatively constant over this period. However, our analysis is limited because, as Table 1 shows, we have relatively little new reviewing activity over this time horizon.

The results of the estimation appear in Table 5. Columns 1 and 2 present estimation results that include differences in average stars and the number of reviews and differences in the fraction of one-star reviews and five-star reviews, respectively, for the whole sample of 2082 books.<sup>17</sup> Columns 4 and 5 present the results for the same specifications for the sample of 275 books that had new reviews at both sites.

The magnitudes on price elasticities in Table 5 are lower than in the cross-sectional specification. The coefficient on changes in prices on Amazon.com is no longer significant. This may be due to relatively little variance in prices over time. In contrast, most of the coefficients on review variables are actually higher in magnitude than in the cross-section sample, even though some are no longer significant.

Qualitatively, most of the results of the previous section are replicated. Thus, an increase in the average star rating on Amazon.com over time results in higher relative sales of the book on Amazon.com over time (one month after the reviews under consideration have been posted). The opposite holds true for changes in average star rating on bn.com. The results for the fraction of five-star reviews and one-star reviews are also consistent with this intuition. Again, we find evidence that one-star reviews have a greater impact than five-star reviews on the same site. As we expected, an increase in the difference in the number of reviews on Amazon.com over time is associated with greater relative sales of the book at Amazon.com over time. The only exception we find is for the difference in number of reviews on bn.com over time. Notably, the coefficient has the "wrong" sign (albeit, it is significant only in the sample of books that had new reviews on each site). However, note that the difference in the change in the number of reviews at Amazon.com and the change in the number of reviews at bn.com continues to be negative. Thus, an increase in the number of reviews at Amazon.com relative to bn.com continues to improve sales at Amazon.com relative to bn.com.

To obtain a sample with more new review activity and as an additional robustness check, we examined changes in reviews and changes in rankings as we did previously, but we examined the change in rankings from May 2003 to May 2004. The new data raise many important issues. First, because the books are all one year older, they are likely to be less popular, and we find that some of the books become unavailable or have missing rankings. Thus, the sample of usable books shrinks to 1636. Second, we discovered that Amazon.com had been active in pruning reviews from the

<sup>16</sup>That is,  $\Delta \ln(\text{PB for book } i) = \ln(\text{PB posted in August 2003 for book } i) - \ln(\text{PB posted in May 2003 for book } i)$ , whereas  $\Delta \ln(\text{number of reviews}^B \text{ on Amazon.com for book } i) = \ln(\text{number of reviews}^B \text{ in July 2003 for book } i) - \ln(\text{number of reviews}^B \text{ in May 2003 for book } i)$ .

<sup>17</sup>If a book has no reviews, we assume that the average star rating of the book is the mean of the books in our sample for that site. In addition, we control for changes from no reviews to reviews, and so forth.



Table 5  
THE EFFECT OF TWO-MONTH CHANGES IN REVIEWS ON CHANGES IN SALES

	1	2	3	4	5	6
Amazon.com $\Delta \ln(\text{price})$	.107 (.232)	.106 (.232)	.096 (.233)	1.591* (.874)	1.419 (.892)	1.228 (.881)
bn.com $\Delta \ln(\text{price})$	-1.426*** (.205)	-1.425*** (.205)	-1.410*** (.205)	-1.500*** (.519)	-1.447*** (.521)	-1.368*** (.522)
Amazon.com $\Delta \ln(\text{number of reviews})$	-.792** (.342)	-.675** (.326)	-.563* (.318)	-1.092 (.955)	-1.026 (.953)	-1.096 (.963)
bn.com $\Delta \ln(\text{number of reviews})$	-.324 (.332)	-.566 (.360)	-.327 (.332)	-1.045* (.604)	-1.146* (.600)	-1.094* (.598)
Amazon.com $\Delta \text{average star rating}$	-.460* (.268)			-1.868 (1.274)		
bn.com $\Delta \text{average star rating}$	.708** (.319)			.832 (.521)		
Amazon.com $\Delta \text{fraction of five-star reviews}$		-.177 (.536)			-3.800 (3.209)	
bn.com $\Delta \text{fraction of five-star reviews}$		1.175** (.587)			1.095 (1.194)	
Amazon.com $\Delta \text{fraction of one-star reviews}$		2.542** (1.283)			4.138 (8.819)	
bn.com $\Delta \text{fraction of one-star reviews}$		-1.057 (1.730)			-3.621 (2.881)	
Amazon.com new five-star reviews			-.003 (.074)			.066 (.198)
bn.com new five-star reviews			.114 (1.730)			-.015 (.189)
Amazon.com new one-star reviews			.061 (.081)			.329** (.153)
bn.com new one-star reviews			-.208 (.231)			-.377 (.310)
Number of observations	2082	2082	2082	275	275	275
Shipping dummies?	Yes	Yes	Yes	Yes	Yes	Yes
R-square	.0391	.0398	.037	.0947	.0951	.0562

\* $p < .10$ .

\*\* $p < .05$ .

\*\*\* $p < .01$ .

Notes: The specification also includes changes in promised shipping times as well as dummies that control for changes from a book having no reviews to having reviews, and so on (in keeping with the cross-sectional specification). For brevity, we omit the coefficients for these variables. The sample in Columns 1–3 is the set of books that were available on both sites in May 2003 and August 2003. The sample in Columns 4–6 is the subsample of the books that had new reviews posted at both sites between May 8 and July 8, 2003. The dependent variable is  $\Delta(\ln[\text{rank}^A] - \ln[\text{rank}^B])$ . If no reviews are present, Amazon.com and bn.com star-ratings variables are set at the meeting star rating for the site. Unreported dummy variables are included that characterize each book for each site into one of the following categories: (1) There were no reviews in May 2003, but there were reviews in the later period; (2) there were no reviews in May 2003, and there were no reviews in the later period; (3) there were reviews in May 2003, but there were no incremental reviews thereafter; and (4) there were reviews in May 2003, and there were incremental reviews thereafter.

sites. Of the 1636 books in the sample, 296 had fewer *total* reviews on Amazon.com in May 2004 than in May 2003. Although these 296 books clearly had reviews removed by Amazon.com, we do not know exactly when these reviews were removed (though we know that we did not have any books experiencing a drop in reviews over the May 2003–August 2003 period).<sup>18</sup>

<sup>18</sup>There are many opportunities to read bloggers' accounts of their reviews being removed by Amazon.com and their hypotheses for why reviews are removed. The removal of reviews does not appear to be strictly from the lower tail. The average number of stars on Amazon.com in May 2003 for books that would have fewer reviews by May 2004 is 4.36, compared with 4.04 for books that would have more or equal reviews in May 2004. Amazon.com states that it removes reviews that are irrelevant. One review that we noticed was removed during this period was a review of a

Table 6 repeats the specifications of Table 5 using the one-year horizon sample. As was the case previously, in Columns 1–3, we constrain the sample to the books that had more reviews in May 2004 than in May 2003. However, recall that Amazon.com appears to have begun removing reviews over the August 2003–May 2004 period, and thus the books that fall into this sample are those for which the number of new reviews exceeds the number of reviews removed.

The results for the average star specification for Amazon.com are entirely insignificant and of the wrong sign in the sample that contains books with no new reviews. Con-

game theory textbook in which the reviewer made extensive reference to the political and religious views of the author.

Table 6  
THE EFFECT OF ONE-YEAR CHANGES IN REVIEWS ON CHANGES IN SALES

	1	2	3	4	5	6
Amazon.com $\Delta \ln(\text{price})$	-.124 (.251)	-.150 (.251)	-.121 (.251)	.695 (.594)	.543 (.596)	.669 (.588)
bn.com $\Delta \ln(\text{price})$	-3.859*** (.287)	-3.859*** (.287)	-3.853*** (.287)	-5.498*** (.586)	-5.444*** (.586)	-5.478*** (.581)
Amazon.com $\Delta \ln(\text{number of reviews})$	-.033 (.032)	-.035 (.032)	-.038 (.032)	-.064 (.161)	-.008 (.160)	.010 (.159)
bn.com $\Delta \ln(\text{number of reviews})$	.026 (.085)	.046 (.087)	.005 (.089)	-.005 (.133)	-.008 (.142)	-.039 (.133)
Amazon.com $\Delta \text{average star rating}$	.020 (.056)			-.189 (.351)		
bn.com $\Delta \text{average star rating}$	.186* (.110)			.405** (.192)		
Amazon.com $\Delta \text{fraction of five-star reviews}$		.108 (.240)			-1.323 (1.772)	
bn.com $\Delta \text{fraction of five-star reviews}$		.024 (.328)			-.036 (.517)	
Amazon.com $\Delta \text{fraction of one-star reviews}$		-.020 (.613)			-1.787 (2.503)	
bn.com $\Delta \text{fraction of one-star reviews}$		-2.049*** (.793)			-2.849** (1.147)	
Amazon.com new five-star reviews			-.171** (.067)			-.336* (.187)
bn.com new five-star reviews			.153* (.087)			.453*** (.176)
Amazon.com new one-star reviews			.052 (.075)			.304** (.136)
bn.com new one-star reviews			-.175 (.125)			-.266* (.162)
Number of observations	1636	1636	1636	459	459	459
Shipping dummies?	Yes	Yes	Yes	Yes	Yes	Yes
R-square	.1223	.1249	.126	.2165	.2222	.236

\* $p < .10$ .

\*\* $p < .05$ .

\*\*\* $p < .01$ .

Notes: The specification also includes changes in promised shipping times as well as dummies that control for changes from a book having no reviews to having reviews, and so on (in keeping with the cross-sectional specification). For brevity, we omit the coefficients for these variables. The sample in Columns 1–3 is the set of books that were available at both sites in May 2003 and May 2004. The sample in Columns 4–6 consists of books that had new reviews posted at both sites between May 2003 and April 2004. The dependent variable is  $\Delta(\ln[\text{rank}^A] - \ln[\text{rank}^B])$ . If no reviews are present, Amazon.com and bn.com star-ratings variables are set at the meeting star rating for the site. Unreported dummy variables are included that characterize each book for each site into one of the following categories: (1) There were no reviews in May 2003, but there were reviews in the later period; (2) there were no reviews in May 2003, and there were no reviews in the later period; (3) there were reviews in May 2003, but there were no incremental reviews thereafter; and (4) there were reviews in May 2003, and there were incremental reviews thereafter.

versely, the coefficient on change in average stars for bn.com is significant and of the expected sign. In the specification that examines the fractions of one- and five-star reviews, we find that the diminished sales created by additional one-star reviews at bn.com remain large and statistically significant.

Because we were concerned that the pruning of reviews by Amazon.com might bias the samples, we attempted a specification that we believed might be somewhat more robust to the pruning exercise. Beginning with the full sample of 1636 books, we coded whether a book had at least one more one-star review than it had before and whether the book had at least one more five-star review than it had before. Thus, in principle, a book could add a one-star

review whether it gained or lost reviews overall, and it could add a five-star review whether it gained or lost reviews overall. Remember that if a person chooses to read all reviews, the reviews are presented from most recent, and the reader must page back to read older reviews. With Amazon.com pruning reviews, it is possible, for example, that an older one-star review was removed, whereas a new one-star review was added, keeping the fraction of one-star reviews the same but moving the one-star review to a more prominent location on the page. Table 6, Columns 1–2 and 4–5, records this book as unchanged, even though it has possibly changed from the reader's perception. Columns 3 and 6 show the results of this specification. Note that the coefficients are all the expected sign; the new five-star coeffi-

cients are significant in both samples, and the new one-star coefficients are significant in the sample of 275 books. Notably, we do not observe that a new one-star review has a greater impact on relative sales than a new five-star review on the same site. This is not simply an artifact of the specification and seems to be sample specific; for comparison, we estimate this specification for the shorter time difference (see Table 5, Columns 3 and 6). In this sample, the impact of the new one-star reviews is greater in magnitude than the impact of new five-star reviews.

### CONCLUSION

We analyze reviewing practices at Amazon.com and bn.com and find that customer reviews tend to be positive at both sites and that they are more detailed at Amazon.com. Our regression estimates suggest that the relative sales of a book across the two sites are related to differences across the sites in the number of reviews for the book and in differences across the sites in the average star ranking of the reviews.

This evidence suggests that customer word of mouth affects consumer purchasing behavior at two Internet retail sites. The notion that customer content affects sales is a prerequisite for differences in customer content quality to have any impact on differences in revenues or profitability across retailers. However, our evidence stops short of showing that retailers profit from providing such content. For example, there is nothing in our evidence that shows that customer reviews do not merely move sales around across books within a site. Because Amazon.com has many more reviewers than rivals and because, on average, its reviews are lengthy and positive, it seems plausible to speculate that the total number of books sold at Amazon.com is higher than it would be without the provision of customer review features. Furthermore, our results show that customers behave "as if" the fit between customer and book is improved by using reviews to screen purchases. An interesting extension to this research would be to examine whether improving a customer's satisfaction with his or her purchases affects subsequent customer loyalty.

There are several worthwhile issues that we leave for further research. For example, we do not explore the review-generating process. This could affect the usefulness of reviews in several important ways. For example, if reviewers respond to previously posted reviews, this may either

decrease or increase the information contained in reviews. On the one hand, an increased dependence on posted reviews could make them less informative. On the other hand, if an unfair or an "incorrect" review prompts a quick reaction, this could increase the overall value of reviews to customers.

### REFERENCES

- Amazon.com (2000), personal e-mail correspondence with customer service agent John Armstrong, (May 15).
- bn.com (2000), personal e-mail correspondence with customer service agent Charlie, (January 14).
- Chevalier, J. and A. Goolsbee (2003a), "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics*, 1 (2), 203–222.
- and ——— (2003b), "Valuing Internet Retailers: Amazon and Barnes and Noble," in *Advances in Applied Microeconomics: Organizing the New Industrial Economy*, Vol. 12, Michael Baye, ed. Amsterdam: Elsevier Science.
- and D. Mayzlin (2003), "The Effect of Word of Mouth on Sales: Online Book Reviews," NBER Working Paper No. 10148.
- Fingar, P., H. Kumar, and T. Sharma (2000), *Enterprise E-Commerce*. Tampa, FL: Meghan-Kiffer Press.
- Godes, D. and D. Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.
- Italie, H. (2001), "Amazon's Bottom 10: Not Exactly Page Turners," *Chicago Sun-Times*, (August 17), 28.
- Mayzlin, D. (2006), "Promotional Chat on the Internet," *Marketing Science*, 25 (2), 157–65.
- McWilliam, G. (2000), "Building Strong Brands Through Online Communities," *MIT Sloan Management Review*, 41 (3), 43–54.
- Morin, R. (2003), "Scholar Invents Fans to Answer His Critics," *The Washington Post*, (February 1), C01.
- Resnick, P. and R. Zeckhauser (2002), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *Advances in Applied Microeconomics: The Economics of the Internet and E-Commerce*, Vol. 11, Michael R. Baye, ed. Amsterdam: Elsevier Science, 127–57.
- Rosenthal, Morris (2005), "Estimating How Many Books Sold by Amazon Rank," (accessed May 1, 2005), [available at <http://www.fonerbooks.com/surfing.htm>].
- Schnapp, M. and T. Allwine (2001), "Mining of Book Data from Amazon.com," paper presented at the UCB/SIMS Web mining conference, (accessed March 24, 2006), [available at <http://www.sims.berkeley.edu:8000/resources/affiliates/workshops/webmining/Slides/ORA.ppt>].