THE EFFECT OF WORD OF MOUTH ON SALES: ONLINE BOOK REVIEWS *

Judith A. Chevalier
judith.chevalier@yale.edu
203-432-3122

Dina Mayzlin
dina.mayzlin@yale.edu
203-436-4262

Yale School of Management
135 Prospect Street
P.O. Box 208200
New Haven, CT 06520-8200
Fax: 203-432-3003

September, 2005

*We would like to thank Pradeep Chintagunta, the Acting Editor, as well as two anonymous reviewers for comments that have greatly improved the paper. We would also like to thank participants at many seminars for helpful comments as well as Sharon Oster, Jackie Luan, David Godes and Jiwoong Shin. We are especially grateful for the comments and encouragement of our late colleague Dick Wittink. Jessie Cheng and Tudor Olteanu provided excellent research assistance. Both authors contributed equally and their names are listed in alphabetical order.

THE EFFECT OF WORD OF MOUTH ON SALES: ONLINE BOOK REVIEWS

We examine the effect of consumer reviews on relative sales of books at Amazon.com and BarnesandNoble.com. We 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, and 3) the impact of 1-star reviews is greater than the impact of 5-star reviews.

Keywords: advertising, word of mouth, source credibility, Internet marketing

1. Introduction

Online user reviews have become an important source of information to consumers, substituting and complementing other forms of firm-to-consumer and offline word of mouth communication about product quality. Consequently, many managers believe that a Website must provide community content in order to build brand loyalty. (See, for example, McWilliams (2000) or Fingar, Kumar, and Sharma (2000)). Despite this widespread belief, to our knowledge, there is no literature documenting that community content plays any role in consumer decision-making. Such a finding, it seems, 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.[1] Second, competing retailers can free-ride on investments in recommender systems; there is nothing to stop a consumer from utilizing the information provided by one website to inform purchases made elsewhere. Third, by providing user reviews, a site cedes control over the information displayed; unfavorable reviews may depress sales. Of course, this may be less of a threat to a retailer that sells many different brands as opposed to a manufacturer. Similarly, since interested parties can freely proliferate favorable reviews for their own products, positive reviews may not be credible and may not function to stimulate sales.[2] Last, online user reviews may not be useful, and may not stimulate sales due to the sample selection bias that

[1] 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.

[2] See Mayzlin (2005) for a theoretical treatment of recommendation systems where firms can anonymously post reviews.

is inherent in an amateur review process. That is, a consumer only chooses to read a book or watch a movie if she perceives that there is a high probability that she will enjoy the experience. In the presence of consumer heterogeneity, this implies that the pool of reviewers will have a positive bias in their evaluation compared to the general population. Thus, positive reviews may simply be discounted by potential buyers.[3]

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 (Amazon) and BN.com, to construct measures of each firm's sales of individual books. Both BN.com and Amazon allow customers to post reviews on the site. Our econometric analysis is designed to answer the 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 quality of) reviews over time for a particular book at one site relative to the other 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 fact that unobserved book characteristics may impact both reviews and sales. By focusing on the *differences* across sites over time, we control for the possibility that taste differences across the populations using the two sites differ in a way that effect both reviews and sales.

[3] In a very different context, Resnick and Zeckhauser (2002) find that 99% of the feedback ratings on ebay.com are positive.

Our findings suggest that reviews tend to be very positive on average, 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 one. We find some evidence that an incremental negative review is more powerful in decreasing book sales than an incremental positive review is in increasing them. Our results on the length of reviews suggests that consumers actually read and respond to written reviews, not merely the average star ranking summary statistic provided by the websites.

The rest of the paper is organized as follows. In Section 2, we describe the data. In Section 3, we describe the methodology. In Section 4, we present our empirical analysis of the effect of customer reviews on product sales. In Section 5, we conclude.

2. Data

Our data consists of individual book characteristics and user review data that were collected from the public Websites of Amazon and BN.com. Our goal was to generate a representative sample of sales. Since we do not have access to the firms' proprietary sales data, we approximate a random sample of sales in the following way: First, we collect a random sample of 3587 books from Global Books in Print released over the 1998-2002 period. However, titles chosen at random will likely have very low sales since 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, since there are few other sources of information on these titles. Thus, in addition, we also collect data on all 2818 titles that appeared in any Publisher's Weekly bestseller list from 1/14/1991 to 11/11/2002, a period ending approximately 6 months prior to our data collection.

We collected data at three time points: for a two-day period in May of 2003, for a two-day period in August of 2003, and for a two-day period in May of 2004. For each book in our sample at each time point, 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 assigned by reviewers (on a scale of 1 to 5 stars, with 5 stars being the best). Most of the books have a promised delivery of 24 hours (96% at Amazon and 88% at BN.com). However, Amazon 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 time points, we extracted detailed characteristics of the most recent 500 reviews of the book posted on the website, including the number of stars assigned and the date the review was posted.[4] 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, with lower sellers assigned higher sequential ranks. We include in our data only books listed as "available" at Amazon.com and Bn.com. Not surprisingly, many of the books drawn randomly from the Books in Print sample were not available for sale at the websites. At the first time point, we find these basic data for 1909 of the Books in Print sample books and for 2261 of the past decade's bestsellers sample.

For each book in our sample, we identify all formats of that book (audio, paperback, hardback, large print, etc). We exclude audio books. Amazon and BN.com provide identical reviews for all of the different formats of a given title. Generally, there is one format that is extremely dominant. Since we do not want the dataset to include duplicate information, we

[4] For the first time point only, we also extracted the full text of the most recent 500 reviews posted on the website.

examine sales and reviews only for the most popular format within a title. We then exclude from our analysis those books for which the most popular format within the title is different at Amazon and BN.com. That is, for example, if the hardcover is the better seller at Amazon and the paperback is the better seller at BN.com, we exclude the book from our sample.[5]

Chevalier and Goolsbee (2003a) report that Amazon 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, 2000). Based on this system then, books that have not been purchased in the past month would not be ranked. Many hundreds of thousands of books, however, have a rank but almost certainly have less than one sale per month. Italie (2001) claims that for these rarely purchased books, Amazon bases the rank on the total sales since Amazon's inception. BN.com claims to update all of its rankings daily (BN.com, 2000).[6] Thus, with the exception of the books that have very high ranks (low sales) on Amazon, the rankings represent a *current* snapshot of sales. However, BN.com only provides 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 does not censor their sales ranks and they appear to range upwards 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. However, books that are relatively

[5] The main results are qualitatively robust to several format selection criteria.

[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 on their ranking system (despite repeated requests).

popular at Amazon and relatively unpopular at BN.com would not appear in the sample, as they have been censored out by BN.com's rank reporting strategy. To address this asymmetry, we remove from our sample those books with ranks above 650,000 at Amazon. More importantly, removing these books serves to remove books for which the Amazon ranks are updated very infrequently. The final sample contains 2387 observations, 1093 of which have reviews posted at both sites at the first (May 2003) time point.

We will be examining differences in sales over the May 2003-August 2003 horizon and over the May 2003-May2004 horizon. As explained above, sales ranks at a particular moment in time represent a snapshot of sales for up to a month. Thus, we decide to undertake a conservative approach to measuring the rank-sales relationship. We examine the relationship between changes in sales over the May 2003-August 2003 time period to changes in reviews over the May 2003 – July 2003 time period. We examine the relationship between changes in sales over the May 2003-May 2004 time period to changes in reviews over the May 2003-April 2004 time period. Because reviews are dated, we can extract the appropriate sample of reviews from the data collected at the August 2003 and May 2004 time points.

We thus rely on sales ranking data in place of 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(sales) and ln(ranks) is approximately linear. Using the methodology proposed there will allow us

to translate sales ranks into sales approximations and thus calibrate the relationship between reviews and sales.[7]

Table 1 presents the summary statistics for the main variables of interest in our data. The number of observations shrinks across time, as books have to be available at Amazon.com and BN.com in the first period to be included in the first period's sample, but have to be available at both sites in both the first period *and* the later period for each of the other two samples. The most striking fact 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 5 stars, and the mean number of stars for any book (with reviews) is greater than 4.

There are, however, 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),[8] (2) Amazon has more reviews than BN.com with BN.com having a much higher fraction of books in our sample that have no reviews at all (54% vs 13% for Amazon), (3) Reviews are slightly more positive on average at BN.com, although, again, they are overwhelmingly positive overall at both sites.

Looking across time, we do not note big changes in pricing or reviewing behavior. Book rankings tend to increase as book popularity falls off. Of course, this leads also to books

[7] This earlier literature approximates that for Amazon, ln(sales) = 9.61 – 0.78ln(rank). For BN.com, following Chevalier and Goolsbee (2003), we scale the relationship down to capture the fact that its sales are 15% of Amazon. In addition, we control for the fact that that there was 24% growth in Amazon in the intervening years. The final relationships are: 9.825-0.78ln(rank) for Amazon and ln(sales)=7.928ln(rank) for BN.com

[8] While BN.com is currently more expensive than Amazon.com, this has not always been true historically. See Chevalier and Goolsbee (2003a), for example.

dropping out of the sample, and thus the summary statistic rankings do not change that 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-August, but review changes from May-July) is very 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 eleven-month reviewing horizon. The data also do not suggest that earlier reviews for a given book are systematically more or less enthusiastic than later ones. The mean change in average star of a book between May 03 and April 04 is within one standard deviation of zero.

Beyond the ratings given by the reviewers, 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 is shown by Godes and Mayzlin (2004). Text analysis programs are very imperfect.[9] However, one 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, at both sites, 1-star and 5-star reviews are much shorter than 2-

[9] Indeed, it has been suggested that we use these data to train text analysis programs. The idea is that a 5-star review must be more enthusiastic than a 4-star or 3-star, and the program can use the reviews to glean patterns that measure levels of enthusiasm.

star, 3-star, and 4-star reviews. Another pattern that emerges is that Amazon reviewers post longer reviews at all star levels than do their peers at BN.com.

3. Model Specification

Consider a book i that is sold on Amazon 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 appropriate because in this case there are scale effects: exogenously, a large number of people look at the "popular" book's page at Amazon/BN.com, and a small number of people look at the "unpopular" book's page at Amazon/BN.com. The fraction of the lookers who go on to buy is plausibly a function of the reviews posted on the site. While log sales would be the ideal dependent variable, we use log rank. Moreover, Schnapp and Allwine (2001), using proprietary data on the sales of a sample of books on Amazon map the relationship between ranks and sales and find that the relationship between log ranks and log sales is very 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 (ν_i), a book-site fixed effect (μ_i^A), as well as other factors. The former is related to factors such as the offline-line promotion, the quality of the book, or the popularity of the author. The latter is related to the fit between the book and the preferences of the customers of the site. That is,

$$\ln(rank^A_i) = \mu_i^A + \nu_i + \alpha^A \ln(P^A_i) + \gamma^A \ln(P^B_i) + X\Gamma^A + S\Pi^A + \varepsilon^A_i \qquad (1)$$

$$\ln(rank^B_i) = \mu_i^B + \nu_i + \alpha^B \ln(P^B_i) + \gamma^B \ln(P^A_i) + X\Gamma^B + S\Pi^B + \varepsilon^B_i \quad (2)$$

where rank denotes the sales rank, the superscripts A and B refer to Amazon and BN.com respectively. P denotes price.[10] X denotes the vector of review variables from both sites: we allow Amazon reviews to affect BN.com's customers and BN.com's reviews to affect Amazon's customers. S is a vector of dummy variables summarizing the shipping times promised by each website 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", etc. For each book, S has a 1 for the promised ship time category at Amazon and a 1 for the promised ship time category for that book at BN.com. We use 4 possible shipping time categories at Amazon, and 3 at BN.com.

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

$$\ln(rank^A_i) - \ln(rank^B_i) = \beta^A \ln(P^A_i) + \beta^B \ln(P^B_i) + X\Gamma + S\Pi + \varepsilon_i \quad (3)$$

However, if there are subtle differences across the two sites (that is, 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,

$$\Delta[\ln(rank^A_i) - \ln(rank^B_i)] = \beta^A \Delta\ln(P^A_i) + \beta^B \Delta \ln(P^B_i) + \Delta X\Gamma + \Delta S\Pi + \varepsilon_i \quad (4)$$

---

[10] We take the log of price in order to estimate the effect of percentage change in price on percentage change in rank.

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

[12] We have some evidence that the two sites' readers and reviewers exhibit very similar preferences. For example, we find that the correlation between log ranks of individual books is very high, 0.825 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 non-fiction received the lowest reviews on both sites. More details can be found at Chevalier and Mayzlin (2003)

The advantage of (3) is that it allows us to utilize more data since many books' reviews do not change over time. In addition, it allows us to estimate the price coefficients since there is not a great amount of variation in prices across time.[13] However, while the differences-in-differences specification 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 site-book specific fixed effects. So, 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, while we briefly present the cross-sectional results, our main focus in the paper is (4).

4. The effect of reviews on sales

4.1 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:

$$\ln(rank^A_i) - \ln(rank^B_i) = \beta^A \ln(P^A_i) + \beta^B \ln(P^B_i) + X\Gamma + S\Pi + \varepsilon_i \qquad (3)$$

Table 3 presents the estimation results for this sample. Column one of Table 3 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 is positive and statistically significant, suggesting that, when prices rise, sales ranks at Amazon become larger, that is, sales fall. The price coefficient is negative for BN.com. This is as expected; recall that the left hand side variable is ln(rank) at Amazon *minus* ln(rank) at BN.com. Again, when prices rise at BN.com, sales ranks become larger, that is, sales fall at BN.com relative to Amazon.

[13] This also allows us to compare our results to previous work.

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. This is consistent with the findings in Chevalier and Goolsbee (2003a) that demand is more elastic at BN.com than Amazon.

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 and the natural log of the total number of reviews at BN. These are set to zero when the number of reviews equals zero. We also include dummies, one that takes the value one when a title at Amazon has no reviews (and zero otherwise) and one that takes the value one when BN.com has no reviews (and zero otherwise). Finally, we also include the average star value of the book's customer reviews at each site in the regression.

As 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, consider a book with four 5-star reviews at both Amazon and BN.com and a rank of 500 at both sites. Now imagine that one of the 5-star reviews at Amazon were changed to a 1-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 would be expected to rise to 601, an estimated decrease in sales of about 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 gets one Amazon review with 1, 2 or 3 stars, its rank on Amazon will rise (sales fall), assuming that its rank on BN.com stays

constant. If, on the other hand, it gets a positive review: 4 or 5 stars, its rank on Amazon will fall (sales rise).

Column 3 focuses on a different way of measuring review valence. In place of average stars, the fraction of reviews that are 1 star reviews and the fraction of reviews that are 5 star reviews are included for each site. As expected, the coefficients suggest that 5 star reviews improve sales and 1-star reviews hurt sales in a statistically significant way at Amazon. The coefficient for 1-star reviews for BN.com is of the expected sign and statistically significant at the 7 percent level. However, the coefficient for 5-star reviews is almost zero but of the "wrong" sign. Nonetheless, it is interesting to note that the 1-star reviews have large coefficients in absolute value, relative to the 5 star reviews, indicating that the relatively rare 1-star reviews carry a lot of weight with consumers. This result also makes sense when one considers the credibility of 1-star and 5-star reviews. After all, the author or other interested party may "hype" his or her own book by publishing glowing reviews on these websites.[14] While the author can post a large number of meaningless 5-star reviews cheaply, he or she cannot prevent others from posting 1-star reviews.[15]

The robustness of these estimates are examined in Columns 4 and 5 of Table 3. 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 presented above. All of the signs of the

[14] For one well-publicized example in economics, see Morin (2003).

[15] One could argue that posting 1 star reviews of competing books could be a reasonable strategy for an author. We acknowledge that this may be true, although it is not at all clear that two books on the same subject, for example, are substitutes rather than complements.

coefficients of interest are as predicted now. The coefficient magnitudes and significance levels for the variables measuring star ratings are somewhat larger than in the full sample.

Finally, we utilize the cross-sectional sample to examine the relationship between review lengths and sales. To do this, we repeat the specifications in Columns 4 and 5 of Table 3, including the natural log of the average length of all of the reviews for each book at each site. The results are presented in Table 4. The coefficient for review length in is positive and statistically significant at Amazon, negative and insignificant at Bn.com. This suggests, controlling 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, more nuanced reviews is in fact harmful to sales. More likely, however, is 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. The ones that read like 4.5 star reviews might on average be longer since they are more likely to be mixed – to mention the negative as well as positive aspects of the book. 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 these 5 star reviews at Amazon is 795 characters for books whose average Amazon star rating is 4 or greater, and is 847 characters for books whose average Amazon rating is less than 4. Similarly, the average review length at Bn.com is 492 for 5 star reviews for a book for which the average rating is 4 or greater, and 675 for 5

star reviews for a book for which the average rating is less than 4. Assuming 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 do seem to suggest that customers read and respond to the review content at each site. However, longer reviews do not necessarily stimulate sales.

4.2 Differences-in-differences Analysis

As discussed earlier, omitted book-site fixed effects could bias the results above. In order to eliminate a possible site-specific fixed effect, we collect review data for May 8 – July 8, 2003, and the ranks, prices, and shipping data as posted on Aug 8, 2003[16] for the sample of 2387 books analyzed in the previous Section. The specification we estimate is:

$$\Delta[\ln(rank^{A}_{i}) - \ln(rank^{B}_{i})] = \beta^{A}\,\Delta\ln(P^{A}_{i}) + \beta^{B}\,\Delta\ln(P^{B}_{i}) + \Delta X\Gamma + \Delta S\Pi + \varepsilon_{i} \qquad (4)$$

Out 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 short time window. However, our analysis is limited by the fact that, as shown in Table 1, we have relatively little new reviewing activity over this time window.

---

[16] That is, $\Delta\ln(P^{B}$ for book i)= $\ln(P^{B}$ posted in August 03 for book i)- $\ln(P^{B}$ posted in May 03 for book i), while $\Delta\ln($Number of reviews^{B} on Amazon for book i)= $\ln($Number of reviews^{B} in July 03 for book i)- $\ln($Number of reviews^{B} in May 03 for book i).
The one month gap between the last review data collected and the rank data collected was to eliminate the possibility that the sales that possibly generated the reviews were included in the dependent measure.

The results of the estimation are presented in Table 5. Columns 1 and 2 present estimation results that include differences in average stars[17] and number of reviews and differences in fraction of 1-stars and 5-stars respectively for the whole sample of 2082 books. Columns 4 and 5 present the results for the same specifications for the sample of 275 books that had new reviews at both sites.

As we can see, the magnitudes on price elasticities in Table 5 are lower than in the cross-sectional specification. In fact, the coefficient on changes in prices on Amazon is no longer significant. This is 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 average star on Amazon over time results in higher relative sales of the book on Amazon over time (one month after the reviews under consideration have been posted). Similarly, the opposite holds for change in average star on BN.com. The results for fraction of 5-stars and 1-stars are also consistent with this intuition. We again find evidence that 1-star reviews have a bigger impact than 5-star reviews on the same site. As expected, an increase in the difference in the number of reviews on Amazon over time is associated with greater relative sales of the book an Amazon over time. The only exception we find is for the difference in number of reviews on BN.com over time. The coefficient is, surprisingly, of the wrong sign (albeit, it is only significant in the sample of books that had new reviews on each site). However, it is important to notice that the difference in the change in the number of reviews

[17] If a book has no reviews, we assume that the average star 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, etc.

at Amazon 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 relative to BN.com continues to improve sales at Amazon relative to BN.com.

In order to obtain a sample with more new review activity and as an additional robustness check, we examine changes in reviews and changes in rankings as above, but examine 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, interestingly, we discovered that Amazon has been active in pruning reviews from the sites. Of the 1636 books in the sample, 296 have fewer *total* reviews on Amazon in May 2004 than in May 2003. While these 296 books clearly have had reviews removed by Amazon, 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 timeframe).[18]

Table 6 repeats the specifications of Table 5 using the one year horizon sample. As before, in Columns 1–3, we constrain the sample to those books that have more reviews in May 2004 than in May 2003. However, recall that Amazon appears to have begun removing reviews over the August 2003-May 2004 period, and that the books which fall into this

[18] As one might imagine, there are many opportunities to read blogger's accounts of their reviews being removed by Amazon and their hypotheses for why reviews are being removed. The removal of reviews does not appear to be strictly from the lower tail. The average number of stars in Amazon in May 2003 for books that would have fewer reviews by May 2004 is 4.36, compared to 4.04 for books that would have greater than or equal reviews in May 2004. Amazon does state that it removes reviews that are irrelevant. One review that we noticed was removed during this time period, was a review of a game theory textbook in which the reviewer made extensive reference to the political and religious views of the author.

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 are entirely insignificant and of the wrong sign in the sample that contains books with no new reviews. On the other hand, the coefficient on change in average star for BN.com is significant and of the expected sign. In the specification examining the fractions of 1s and 5s, we find that the diminishment of sales created by additional 1-star reviews at BN.com remains large and statistically significant.

Because we have some concerns that the pruning of reviews by Amazon might bias the samples, we attempt a specification that we think might be somewhat more robust to the pruning exercise. Starting with the full sample of 1636, we code whether a book has at least one more 1-star review than it had before and whether the book has at least one more 5-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 5-star review whether it gained or lost reviews overall. Remember that if one chooses to see all reviews, the reviews are presented from most recent, with the reader paging back to see older and older reviews. With Amazon having started pruning reviews, it is possible, for example, that an older 1-star review was removed, while a new 1-star review was added, keeping the fraction of 1 stars the same, but moving the 1-star review to a more prominent location on the page. Columns 1-2 and 4-5 of Table 6 will record this book as unchanged, even though it has possibly changed in the reader's perceptions. Columns 3 and 6 of Table 6 show the results of this specification. Note that the coefficients are all of the expected sign; the new 5-star coefficients are significant in both samples, and the new 1-star coefficients are significant in the sample of

275 books. Interestingly, we do not observe that a new 1-star review has a bigger impact on relative sales than a new 5-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 that sample, the impact of the new 1-star reviews is bigger in magnitude than the impact of new 5-star reviews.

7. Conclusion

We analyze reviewing practices at Amazon and BN.com. We find that customer reviews tend to be very positive at both sites and that they are more detailed at Amazon. Our regression estimates suggest that the relative sales of a book across the two sites is 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. That customer content impacts sales is certainly a prerequisite for differences in customer content quality to have any impact on differences in revenues or profitability across retailers. Our evidence however, stops short of showing that the retailer profits 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. Since Amazon has many more reviewers than rivals, its reviews are on average quite lengthy, and its reviews are on average quite positive, it seems plausible to at least speculate that the total number of books sold at Amazon is higher than it would have been absent the provision of customer review features. Further, and more interestingly, our results show that customers certainly behave *as if* the fit between customer and book is improved by using

reviews to screen purchases. One 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 a number of interesting questions that we leave for future research. For example, we do not at all explore the review-generating process. This could impact the usefulness of reviews in several important ways. For example, if reviewers respond to earlier posted reviews, this may either decrease or increase the information contained in reviews. On one hand, an increased dependence in posted reviews would seem to worsen its informativeness. On the other hand, if an unfair or an "incorrect" review prompts a quick reaction, this may overall increase the overall value of reviews to customers.

Tables

Table 1: Summary data.

The sample is 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 is the same as the most popular format of the book at BN.com. Means are primary data entry, with standard deviations in parentheses below.

	May-03		Aug-03		May-04	
	Amazon	BN.com	Amazon	BN.com	Amazon	BN.com
Price	13.97	15.50	13.85	15.2	13.56	15.22
	(14.41)	(14.75)	(14.84)	(15.28)	(15.12)	(15.79)
Ranking	129799.00	121061.00	134303	122377	123112	137402
	(169363)	(156903)	(166575)	(152466)	(152349)	(166939)
Number of reviews/book	60.99	12.79	59.79[19]	13.11	68.31	14.15
	(180.40)	(44.55)	(183.70)	(46.70)	(205.42)	(42.30)
Average stars	4.14	4.45	4.13	4.16	4.06	4.43
	(0.70)	(0.57)	(0.71)	(0.62)	(0.70)	(0.58)
Fraction 1s	0.07	0.03	0.07	0.03	0.08	0.04
	(0.12)	(0.08)	(0.12)	(0.08)	(0.12)	(0.09)
Fraction 5s	0.57	0.67	0.57	0.67	0.51	0.66
	(0.29)	(0.26)	(0.29)	(0.26)	(0.29)	(0.26)
Incremental reviews per book			1.82	0.53	10.56	1.85
			(5.49)	(2.27)	(67.59)	(18.72)
Change in average stars			-0.010	-0.010	-0.006	-0.015
			(0.16)	(0.14)	(0.50)	(0.27)
Fraction books with no reviews	0.13	0.54	0.12	0.54	0.17[20]	0.49
Number of obs	2387	2387	2082	2082	1636	1636

[19] Note that this number is slightly lower than the average number of reviews per book in May 03 Amazon sample. This is not due to a loss of reviews over this time period, but is due 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 03 and thus were not included in the sample later.

[20] Note that the fraction of books with no reviews on Amazon goes up in this time period. This is at least partially due to the pruning of reviews by Amazon that is discussed in the paper.

Table 2: Review length and star distribution for the May '03 sample. The sample includes all books with reviews.

	Amazon		BN.com	
	frequency	char length	frequency	char length
1 star reviews	8.97	765	3.44	558
2 star reviews	7.53	916	4.07	599
3 star reviews	10.56	997	6.00	566
4 star reviews	19.89	949	19.27	577
5 star reviews	53.05	812	67.22	508
Overall		854		529

Table 3: The effect of reviews on sales. In columns 1, 2, 3 the sample is the May 2003 sample, as in table 1. In Columns 4 and 5, the sample is those books that have at least one review on both sites in May 2003. The dependent variable is the difference between the log ranking of the book at Amazon and the log sales ranking of the book at BN.com. That is, the dependent variable is $Ln(rank^A) - Ln(rank^B)$.

	(1)		(2)		(3)		(4)		(5)	
Amazon ln(price)	1.556	***	1.545	***	1.532	***	2.147	***	2.148	
	(0.159)		(0.155)		(0.156)		(0.324)		(0.328)	
BN ln(price)	-1.801	***	-1.837	***	-1.826	***	-2.67	***	-2.58	
	(0.148)		(0.144)		(0.145)		(0.280)		(0.282)	
Amazon ln(no. of reviews)			-0.215	***	-0.205	***	-0.403	***	-0.373	
			(0.024)		(0.023)		(0.050)		(0.050)	
BN ln(no. of reviews)			0.131	***	0.13	***	0.259	***	0.242	
			(0.033)		(0.033)		(0.052)		(0.052)	
Amazon no reviews dummy			-0.574	***	0.075	***				
			(0.187)		(0.109)					
BN no reviews dummy			-0.154		-0.354	**				
			(0.100)		(0.131)					
Amazon average star rating			-0.184	***			-0.418	***		
			(0.038)				(0.079)			
BN average star rating			0.024				0.145	*		
			(0.017)				(0.088)			
Amazon fraction reviews 5 star					-0.256	***			-0.704	***
					(0.100)				(0.235)	
BN fraction reviews 5 star					-0.147				0.061	
					(0.149)				(0.188)	
Amazon fraction reviews 1 star					0.483	**			1.15	**
					(0.255)				(0.506)	
BN fraction reviews 1 star					-0.836	*			-0.94	*
					(0.467)				(0.566)	
No. observations	2387		2387		2387		1087		1087	
includes shipping dummies?	y		y		y		y		y	
R-squared	0.086		0.138		0.136		0.216		0.203	
*** p < 0.01										
** p < 0.05										
* p < 0.10										

Table 4: The Effect of review length on book market shares.
The sample is as in Columns 4 and 5 of Table 3.
Dependent variable ln(rank) at Amazon minus ln(rank) at Bn.com.

	(1)	(2)
Amazon ln(price)	2.127***	2.093***
	(0.325)	(0.326)
BN ln(price)	-2.661***	-2.634***
	(0.279)	(0.280)
Amazon ln(no. of reviews)	-0.415***	-0.411***
	(0.0501)	(0.0503)
BN ln(no. of reviews)	0.267***	0.267***
	(0.0518)	(0.052)
Amazon average star rating	-0.405***	
	(0.0794)	
BN average star rating	0.138*	
	(0.0878)	
Amazon fraction reviews 5 star		-0.441*
		(0.242)
BN fraction reviews 5 star		0.083
		(0.188)
Amazon fraction reviews 1 star		1.550***
		(0.511)
BN fraction reviews 1 star		-1.020*
		(0.563)
Amazon ln(average rev length)	0.570***	0.598***
	(0.146)	(0.151)
Bn ln(average rev length)	-0.049	-0.052
	(0.0917)	(0.0920)
No. observations	1087	1087
Includes shipping dummies?	Y	Y
R-squared	0.217	0.216

Table 5:[21] The effect of change in reviews on change in sales.
The sample in (1) - (3) is the same as in Table 4, with the additional restriction that the books had to be available in both times. The sample in (4) - (6) consists of books that, in addition, had non-zero posted reviews at both sites in May 8 – July 8, 2003.
Dependent variable is Δ(ln(rank) at Amazon minus ln(rank) at Bn.com).

	(1)		(2)		(3)		(4)		(5)		(6)	
Amzn Δln(price)	0.107		0.106		0.096		1.591	*	1.419		1.228	
	(0.232)		(0.232)		(0.233)		(0.874)		(0.892)		(0.881)	
BN Δln(price)	-1.426	***	-1.425	***	-1.410	***	-1.500	***	-1.447	***	-1.368	***
	(0.205)		(0.205)		(0.205)		(0.519)		(0.521)		(0.522)	
Amzn Δln(no. of revs)	-0.792	**	-0.675	**	-0.563	*	-1.092		-1.026		-1.096	
	(0.342)		(0.326)		(0.318)		(0.955)		(0.953)		(0.963)	
BN Δln(no. of revs)	-0.324		-0.566		-0.327		-1.045	*	-1.146	*	-1.094	*
	(0.332)		(0.360)		(0.332)		(0.604)		(0.600)		(0.598)	
Amzn Δavg star rating	-0.460	*					-1.868					
	(0.268)						(1.274)					
BN Δavg star rating	0.708	**					0.832					
	(0.319)						(0.521)					
Amzn Δfrac revs 5 star			-0.177						-3.800			
			(0.536)						(3.209)			
BN Δfrac revs 5 star			1.175	**					1.095			
			(0.587)						(1.194)			
Amzn Δfrac revs 1 star			2.542	**					4.138			
			(1.283)						(8.819)			
BN Δfrac revs 1 star			-1.057						-3.621			
			(1.730)						(2.881)			
Amzn new 5 star					-0.003						0.066	
					(0.074)						(0.198)	
BN new 5 star					0.114						-0.015	
					(1.730)						(0.189)	
Amzn new 1 star					0.061						0.329	**
					(0.081)						(0.153)	
BN new 1 star					-0.208						-0.377	
					(0.231)						(0.310)	
No. observations	2082		2082		2082		275		275		275	
shipping dummies?	y		y		y		y		y		y	
R-squared	0.0391		0.0398		0.037		0.0947		0.0951		0.0562	

[21] 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, etc. (in keeping with the cross-sectional specification). For brevity, the coefficients for these variables are omitted.

Table 6[22]: The effect of change in reviews on change in sales.
The sample in (1) - (3) is the same as in Table 4, with the additional restriction that the books had to be available in both times. The sample in (4) - (6) consists of books that, in addition, had non-zero posted reviews at both sites in May 2003 – May 2004.

Dependent variable is Δ(ln(rank) at Amazon minus ln(rank) at Bn.com).

	(1)		(2)		(3)		(4)		(5)		(6)	
Amzn Δln(price)	-0.124		-0.150		-0.121		0.695		0.543		0.669	
	(0.251)		(0.251)		(0.251)		(0.594)		(0.596)		(0.588)	
BN Δln(price)	-3.859	***	-3.859	***	-3.853	***	-5.498	***	-5.444	***	-5.478	***
	(0.287)		(0.287)		(0.287)		(0.586)		(0.586)		(0.581)	
Amzn Δln(no. of revs)	-0.033		-0.035		-0.038		-0.064		-0.008		0.010	
	(0.032)		(0.032)		(0.032)		(0.161)		(0.160)		(0.159)	
BN Δln(no. of revs)	0.026		0.046		0.005		-0.005		-0.008		-0.039	
	(0.085)		(0.087)		(0.089)		(0.133)		(0.142)		(0.133)	
Amzn Δavg star rating	0.020						-0.189					
	(0.056)						(0.351)					
BN Δavg star rating	0.186	*					0.405	**				
	(0.110)						(0.192)					
Amzn Δfrac revs 5 star			0.108						-1.323			
			(0.240)						(1.772)			
BN Δfrac revs 5 star			0.024						-0.036			
			(0.328)						(0.517)			
Amzn Δfrac revs 1 star			-0.020						-1.787			
			(0.613)						(2.503)			
BN Δfrac revs 1 star			-2.049	***					-2.849	**		
			(0.793)						(1.147)			
Amzn new 5 star					-0.171	**					-0.336	*
					(0.067)						(0.187)	
BN new 5 star					0.153	*					0.453	***
					(0.087)						(0.176)	
Amzn new 1 star					0.052						0.304	**
					(0.075)						(0.136)	
BN new 1 star					-0.175						-0.266	*
					(0.125)						(0.162)	
No. observations	1636		1636		1636		459		459		459	
shipping dummies?	y		y		y		y		y		y	
R-squared	0.1223		0.1249		0.126		0.2165		0.2222		0.236	

[22] 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, etc. (in keeping with the cross-sectional specification). For brevity, the coefficients for these variables are omitted.

References

1. Amazon, e-mail correspondence with customer service agent John Armstrong, May 15, 2000.

2. Avery, C.; Resnick, P.; Zeckhauser, R. (1999) "The Market for Evaluations," *American Economic Review*. 89 (June), 564-84.

3. BN.com, e-mail correspondence with customer service agent Charlie, January 14, 2000.

4. Brynjolffson, E. and Smith, M. (2000) "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," Management Science 46, 563-585.

5. Chevalier, J. and Goolsbee, A. (2003a) "Measuring prices and price competition online: Amazon.com and BN.comandNoble.com," *Quantitative Marketing and Economics* I(2), forthcoming.

6. Chevalier, J. and Goolsbee, A. (2003) "Valuing Internet Retailers: Amazon.com and BN.comandNoble.com," Yale School of Management working paper.

7. Chevalier, J. and Mayzlin, D. (2003) "The Effect of Word of Mouth on Sales: Online Book Reviews," NBER Working Paper 10148.

8. Clay, K., Krishnan, R. and Wolff, E. (2001) "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *The Journal of Industrial Economics* 49, 521-540.

9. Coleman, J. S., E. Katz, and Menzel, H. (1966). "Medical Innovation: A Diffusion Study," Indianapolis, Indiana: Bobbs-Merrill.

10. Eliashberg, J. and Shugan, S. "Film critics: Influencers or predictors?" *Journal of Marketing*, Vol. 61, No. 2 (April 1997), 68-78.

11. Ellison, G.; Fudenberg, D. (1995) "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*. 110 (1), 93-125.

12. Fingar, P.; Kumar, H.; Sharma, T. (2000) *Enterprise E-Commerce*, Meghan-Kiffer Press.

13. Foster, A. and Rosenzweig, M. (1995, December). "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*. 103(6), 1176–1210.

14. Godes, D. and Mayzlin, D. (2004) "Using Online Conversations to Study Word of mouth Communication," *Marketing Science*, 23 (4), 545 – 560.

15. Italie, H. (2001), "Amazon's Bottom 10: Not Exactly Page Turners," *Chicago Sun-Times*, 8/17/ 2001.

16. Katz, E. and Lazarsfeld, P. (1955). Personal Influence. Glencoe, IL: Free Press.

17. Mayzlin, D. (2005) "Promotional Chat on the Internet," forthcoming in *Marketing Science*.

18. McWilliam, G. (2000) "Building Strong Brands through Online Communities," *MIT Sloan Management Review*, 41 (3), 43-54.

19. Morin, R. (2003) "Scholar Invents Fans to Answer his Critics," *Washington Post*, 2/1/2003, page C01.

20. Resnick, P. and Zeckhauser, R. (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," *The Economics of the Internet and E-Commerce*. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics. Amsterdam, Elsevier Science.

21. Rosenthal, M. (2005) "Estimating How Many Books Sold by Amazon Rank", http://www.fonerbooks.com/surfing.htm.

22. Schnapp, M. and Allwine, T. (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

23. Van den Bulte, C. and G. Lilien (2001). "Medical Innovation revisited: Social contagion versus marketing effort," *American Journal of Sociology* 106 (5), 1409–35.