Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings*

Imke Reimers
Northeastern University

Joel Waldfogel
University of Minnesota, NBER, ZEW

February 2, 2020

Abstract

Digitization has led to a large number of new creative products, straining the capacity of professional critics and consumers. Yet, the digitization of retailing has also delivered new crowd-based sources of pre-purchase information. We compare the relative impacts of professional critics and crowd-based Amazon star ratings on consumer welfare in book publishing. Using various fixed effects and discontinuity-based empirical strategies, we estimate their causal impacts on sales. We use these causal estimates to calibrate a structural demand model. The aggregate effect of star ratings on consumer surplus is roughly 15 times the effect of traditional review outlets.

When choosing among experience goods, consumers benefit from guidance prior to purchase. Traditionally, professional critics – such as product reviewers in prominent media outlets – played important roles in providing this guidance.[1] One of digitization's many impacts has been a sharp increase in the number of new creative products. While the number of new products has always exceeded the capacity of professionals to review them, this gap has only grown with digitization.[2] Crowd-based ratings – such as Amazon stars – on the other hand, are available for all products, raising the possibility that another facet of digitization, ubiquitous crowd ratings, can provide information that allows the realization of welfare gains from new products.

These considerations raise the question of how the new crowd-based pre-purchase information made available by digitization affects purchase behavior and, by extension, welfare. To address this, we ask the following specific questions. First, do professional reviews and crowd ratings have causal impacts on demand; and if so, how large are these impacts? Second, are crowd-based ratings available for a different mix of books than those reviewed by professional critics? Third, how do the two pre-purchase information institutions – professional reviews and crowd ratings – affect the welfare of consumers? And finally, we ask two questions about the potential substitutability of reviews and crowd ratings: Do crowd-based ratings function as substitutes or complements for a title's professional reviews? And over time, has the growth of crowd-based rating systems reduced the influence of professional reviews?

We address these questions in the market for books. Books provide an auspicious study context for a few reasons. First, books are experience goods, so that pre-purchase information is potentially useful. Second, the number of professional reviews, and particularly the number appearing in highly visible outlets, is relatively small and therefore feasible to observe and quantify. Third, and perhaps most important, we have high-frequency data on book demand

---

[1]See Deutschman (2004); Pompeo (2017), or Martin (2011) for descriptions of various professional critics and their influence on product markets.

[2]See Waldfogel (2017) for evidence on the growth in new products. In 2014 New York Times film critic Manohla Dargis implored the film industry to make fewer movies. See Dargis (2014).

at Amazon – which accounts for about 45 percent of the US physical book market – that helps identify causal relationships.[3] We have daily measures of Amazon sales ranks and their crowd-based star ratings, for 7,947 titles (appearing in 10,549 editions) during 2018, for three English-language Amazon sales domains (the US, Canada, and the UK).

Reviews and star ratings are inherently endogenous, as raters and reviewers decide whether and when to give feedback, in addition to what they say. More appealing books sell more and receive more positive feedback. Our high-frequency data from multiple platforms allow us to deal with this endogeneity using two strategies, one for reviews and one for star ratings. We treat the appearance of a professional review as a discontinuous jump in attention delivered to the title, and we look for a corresponding jump in our daily sales measure. We measure the impacts of star ratings with a cross-platform longitudinal comparison in the spirit of Chevalier and Mayzlin (2006), employing both book fixed effects and cross-platform intertemporal comparisons.

Our descriptive analysis gives us causal evidence on the links between pre-purchase information – reviews and ratings – and sales ranks. Quantification of the importance of these two mechanisms requires a framework for welfare analysis, which, in turn, requires two translational steps. First, we transform effects of pre-purchase information on sales ranks into effects on quantities, allowing the calculation of the elasticities of quantity sold with respect to the Amazon price and the star rating, as well as the percentage impacts of professional reviews on sales. Second, we use those elasticities to calibrate nested logit models of demand that facilitate welfare analysis. Our measured welfare effects of ratings and reviews allow the consumers' ex ante choice utilities – which may be based on limited information – to differ from their ex post consumption utilities.

We have four broad findings. First, professional review outlets, notably the New York Times, have clear impacts on sales. In the five days following a New York Times review, a book's estimated sales improve by 78 percent, on average, with slightly larger effects for

---

[3]See https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/78929-print-unit-sales-increased-1-3-in-2018.html.

more positive reviews. Over the entire year, a New York Times review raises sales by 4.6 percent. Second, the crowd also has clear effects on sales: The elasticity of sales with respect to Amazon stars averages about 0.75, and it is larger when the stars are based on more underlying ratings. Third, professional reviews and star ratings have similar per-title impacts on consumer surplus; but in the aggregate, professional reviews raise consumer surplus by just under $3 million while the aggregate effect of Amazon star ratings ($41 million) is roughly 15 times larger. Fourth, we do not find evidence of substitution between reviews and star ratings at either the book level or in the aggregate. The effect of a star rises after the appearance of a professional review; and the effects of New York Times reviews, in a supplementary analysis covering weekly sales from 2004 to 2018, have not waned with the growth of digitization.

We conclude that digitization has delivered not only a proliferation of new products but also new information mechanisms that add substantially to the value of the pre-purchase information available to consumers from traditional review sources. These crowd-based reviews provide pre-purchase information on all products, including those neglected by professional critics, and they do so without undermining effects of professional critics on the books and genres professionals do cover.

The paper proceeds in seven sections. Section 1 provides background on the book market, the evolution of the information environment with digitization, and a discussion of the existing literature. Section 2 presents a simple theory of choice with and without pre-purchase product information, which organizes our descriptive and welfare analyses. We also contrast the roles played by professional reviews and crowd ratings in affecting buying behavior. Section 3 describes our data on Amazon sales ranks, star ratings, and prices, as well as reviews in major newspapers. Section 4 presents our empirical strategies for measuring causal impacts of professional reviews and star ratings on sales ranks. We also present estimated effects on log sales ranks, as well as translations of those estimated effects into effects on quantities sold. Section 5 then turns to welfare analysis. Using structural demand models calibrated to

our causal quantity estimates, we measure the respective welfare gains arising from Amazon star ratings and professional reviews. Section 6 asks whether the effect of New York Times reviews on sales changed between 2004 and 2018. Section 7 concludes.

1 Background

1.1 The US Product and Information Environment for Books

In 2000, roughly 80,000 fiction and non-fiction titles were released in the US, and the number of new titles released annually has grown sharply since then. In 2012, when 100,000 new US titles appeared in hardback form, the number of new US ebook titles was 280,000.[4] This figure, while impressive, only counts the titles with ISBNs ("international standard book number"), which many self-published titles lack. Clearly, there has been substantial growth in the number of new book titles released in the US. Large physical bookstores only carry roughly 200,000 new and old titles, so only a small fraction of new titles have traditionally been marketed directly to consumers (Greenfield, 2012). Even before digitization, product discovery was a significant challenge; the challenge has grown substantially since.[5]

1.2 Professional Reviews

There is a two-part professional reviewing ecosystem that supports retailer, library, and consumer discovery of new products. One part consists of reviews targeted at libraries and bookstores, from outlets such as Publishers Weekly, Library Journal, and Kirkus. These "B2B" outlets review relatively large numbers of titles – although a small share of releases – but have rather limited audiences. In 2018, they each reviewed about 3,500 to 7,500 books, and their respective sites attracted no more than 2.15 million visits in December

---

[4]These figures are based on queries of the Bowker Books in Print database for numbers of English-language hardback and ebook titles published in the US.

[5]See Waldfogel and Reimers (2015) for additional data on the growth in new books since digitization.

2018, according to Similarweb.[6]

The other, consumer-facing part of the reviewing environment consists mainly of reviews in daily newspapers, including the New York Times, the Wall Street Journal, the Washington Post, the Los Angeles Times, the Boston Globe, and the Chicago Tribune, all of which we include in our sample. We describe the sample in detail below, but here we note that the New York Times has far more reviews than the others. Our sample for 2018 includes 1,114 titles reviewed by the New York Times, compared with between 57 and 149 for each of the others. Of the titles reviewed in any of the other papers, just over half were also reviewed in the New York Times. Newspaper websites have far more traffic and visibility than the sites for B2B book review outlets. For example, the Washington Post had 120.5 million monthly visitors in December 2018, while the New York Times had 302.5 million, according to Similarweb. Measured by both volume of reviews and visibility to consumers, the New York Times is the preeminent US book review outlet.[7]

1.3 Crowd-based Star Ratings at Amazon

Amazon allows users to review and rate books on a five-point scale, and Amazon aggregates this user feedback into star ratings for each book. A few features of the ratings system are noteworthy. First, in contrast to professional reviews, which are available for a small share of titles, crowd ratings are available for all of them. Second, as users leave ratings, Amazon aggregates these individuals' ratings into an overall rating, which they report to a tenth of a star, although the aggregation is not a simple averaging.[8] Third, leaving ratings is common,

---

[6]We get a rough count of the number of titles reviewed during 2018 by querying Bowker's Books in Print, which contains indicators for whether a book was reviewed by each of a number of major outlets.

[7]We ignore magazines because they have small reach or review few books. For example, the New Yorker, which typically provides one long and four short reviews per weekly issue, had 14.9 million site visitors according to Similarweb, about a tenth of the volume at the Washington Post, which we show below to have a negligible effect on sales. While Oprah's book club has been documented to have large effects (Garthwaite, 2014), the club has reviewed only four books per year on average (`https://www.oprahmag.com/entertainment/books/g23067476/oprah-book-club-list/`).

[8]Rather, "Amazon calculates a product's star ratings based on a machine-learned model instead of a simple average. ...These models take into account factors such as how recent the rating or review is and verified purchase status. They use multiple criteria that establish the authenticity of the feedback. The system

and consumers can easily observe the number of underlying ratings on which the visible star rating is based. While all books start with no ratings, the average title in our sample, described in more detail below, has 326 underlying ratings by the end of 2018. Fourth, the star rating for a particular book differs across Amazon's country platforms.[9]

1.4   Existing Literature

Our study is related to three existing literatures. First, it is related to work measuring the impact of professional reviews on product sales. Reinstein and Snyder (2005), Sorensen (2007), Berger et al. (2010) and Garthwaite (2014) provide four examples of studies employing careful empirical strategies to document impacts of professional reviews on movie and book sales. Existing studies of reviews and book sales document causal impacts using weekly sales data. We are able to build on this work using higher-frequency, daily data for a large sample of books.

Second, our study is related to existing work on the impact of word of mouth reviews on sales. Prominent examples include Chevalier and Mayzlin (2006); Luca (2016); Duan et al. (2008); Forman et al. (2008); Helmers et al. (2019), and Senecal and Nantel (2004). Chevalier and Mayzlin (2006) makes use of a cross-platform comparison of books' sales ranks and star ratings to measure impacts of "word of mouth," in the form of star ratings, on sales. Finally, because our structural welfare analysis makes the distinction between ex ante "decision utility" and ex post experienced utility, we build on a third literature (Jin and Sorensen, 2006, Allcott, 2011, and Train, 2015).

---

continues to learn and improve over time" (`https://www.amazon.com/gp/help/customer/display.html`).

[9]Digitization has also fostered growth in amateurs who distribute their book reviews at Goodreads. Because these reviewers appear not to have effects on book sales; we relegate their discussion to Appendix section A.

2 Theory: Information, Purchase, and Welfare

2.1 The Roles of Ratings and Reviews in Product Purchase

While both reviews and ratings are pre-purchase information, consumers interact with them in different ways; and they may have different effects on purchase. A subset of books is reviewed by professionals, and their reviews are delivered to consumers as newspaper articles. Consumers seeing these reviews may have no prior familiarity with a book being reviewed. A review may therefore both create awareness of a book, as well as delivering information about its quality. Both awareness and quality information can change the tendency for a consumer to purchase a product.

Consumers interact differently with Amazon star ratings. Consumers encounter these ratings when shopping for particular books. Rather than alerting consumers to a book's existence, ratings provide quality assessments for those consumers already considering those titles.

While both reviews and ratings convey quality information and are therefore in some broad sense substitutes, they may also act as complements for particular titles. It is possible, for example, that reviews induce consumers to shop for a particular title and, by driving more traffic to the book's page, raise the star rating's effect on purchase. It is also possible that a review, by virtue of its credibility with readers, renders the star rating ineffective at promoting purchase. Therefore, whether the relationship between reviews and ratings is one of substitutability or complementarity is an empirical question. We explore this below at the level of particular titles by asking how the effects of star ratings change after a book has received a professional review. In section 6, we also ask the broader question of whether the effect of professional reviews has changed with the growing availability of star ratings over the past decade and a half.

2.2 Information and Purchase

Regardless of their mechanisms, reviews and ratings provide information that can affect consumers' tendencies to purchase products. To be concrete – and to put this in a framework we revisit to below – suppose a consumer i has the following utility function for a product j when reviews exist:

$$u_{ij} = u(R_j, p_j; x_j)$$

In this setup, R_j is the pre-purchase product information (rating or review) on product j, p_j is the product's price, and x_j contains other observables on product j. Because pre-purchase information exists, R_j is both a measure, and the pre-purchase indicator, of quality.

If reviews and ratings did not exist, then consumers might instead form predictions of quality based on characteristics of the product, which we summarize in this setup as a predicted rating, $\hat{R}_j$. Expected utility absent the reviews would then be

$$u_{ij} = u(\hat{R}_j, p_j; x_j).$$

A surprisingly positive rating – when a product is better than expected so that $R_j > \hat{R}_j$ – could increase its consumption relative to its consumption in their absence, and vice versa. Whether this would happen, of course, depends on the causal impact of review information on purchase (and therefore, we infer, utility). Hence, our main empirical task below is to measure the causal impact of reviews and ratings on purchase.

2.3 Pre-Purchase Information and Welfare

Assessing effects of pre-purchase information on welfare requires a distinction between expected ex ante utility and experienced ex post utility. For this, we follow studies such as Jin and Sorensen (2006), Allcott (2011), and Train (2015).

Suppose that consumers were uninformed prior to purchase and, in particular, that they

believed the product's quality to be lower than its true quality ($\hat{R}_j < R_j$ in the above example). Then their ex ante aggregate demand curve for the title would be given by the dashed curve in Figure 1. They would choose Q_1 units, and at purchase they would expect consumer surplus equal to region A. Upon consumption, however, they would perceive the product's true value, so that the ex post experienced consumer surplus would be regions $A + B$. Had they been informed prior to purchase, they would have chosen Q^* units and would have experienced their ex ante CS – regions $A + B + C$ – as ex post consumer surplus. Therefore, the value of access to this pre-purchase review information is the difference between these surplus regions, or C.

There is an analogous case, in which consumers believe the product is better than it actually is ($\hat{R}_j > R_j$) and consume Q_2 units. While the consumers expected even more prior to purchase, their experienced consumer surplus is regions $A + B + C$ less region D. If they had access to information prior to purchase, they would have consumed Q^*, generating consumer surplus of $A + B + C$. Hence, the value of information to these consumers is region D. Generically, the welfare gain from having pre-purchase information arises from a "triangle" associated with consuming either too much or too little of the product in the absence of having pre-purchase information. The base of this triangle is the amount by which quantity deviates from the informed quantity, and its height is determined by the shape of the demand curve for the product. To say this another way, the effect of pre-purchase information on consumption – and the consequent difference between Q^* and chosen consumption – determines the size of the welfare effect.

The empirical welfare analysis requires characterizations of the environment without professional reviews and star ratings. We model the absence of these types of pre-purchase information differently. Professional reviews are available for a small fraction of titles. While reviews can in principle be positive or negative, it turns out empirically – both in our data and in previous work (Berger et al., 2010) – that reviews tend to have positive impacts on sales. Moreover, while we show that more positive reviews generate higher sales, the

difference between the impacts of positive and less positive reviews is small. As a result, it is useful to view reviews – unlike star ratings – as binary. Either a book is reviewed, or it is not. For the counterfactual case eliminating reviews, we model all books as not having been reviewed. We handle crowd ratings differently. In the absence of ratings, we assume consumers would expect title j to have quality $\hat{R}_j$. We discuss the associated prediction of this counterfactually assumed quality in section 5.

3 Data

3.1 Data Set Construction

The ideal dataset for addressing our questions would be a high-frequency panel on prices and quantities, as well as ratings and review information for every book published over some period, or at least a representative sample of all titles, including those reviewed by major outlets. Our data resemble the ideal in some respects but also have some features that require adaptation.

In what follows, we first explain which book titles are included in the study. Second, we describe how we obtain each professional outlet's review timing. Third, we explain how we obtain lists of ISBNs for particular editions of each title, which we use to get daily Amazon data on prices, sales ranks, and star ratings. Finally, we describe the features of the data that require adaptation.

We create a list of books that reflects what sells by starting with the most comprehensive publicly available bestseller list, the USA Today weekly top 150.[10] During 2018, this list includes 2,581 distinct titles (4,290 editions). We supplement this list with all books reviewed in the New York Times and the other major review outlets during 2018 – 1,403 titles (1,973 editions) – as well as a list of books of interest to lay readers outside of the right tail of the sales distribution. These are the 2,386 titles (3,500 editions) published in 2018 and reviewed

---

[10]See https://www.usatoday.com/entertainment/books/best-selling/.

in the same year by widely followed users of the site Goodreads.[11] Given the overlap across lists, the grand list includes 7,947 distinct titles (10,549 editions). Most of these are published during 2018, but some are published earlier.

We obtain review dates for all books reviewed in the New York Times by directly searching at the newspaper's website. For the other newspapers (the Boston Globe, the Chicago Tribune, the Los Angeles Times, the Wall Street Journal, and the Washington Post), we use the Bowker Books in Print directory. We obtain lists of hardcover editions published in 2018 and reviewed by the newspaper, and we then find the reviews written in 2018 using Google searches of, say, "Chicago Tribune book review [author title]." For books reviewed by the New York Times, we also have a measure of whether the title was among those more favorably reviewed, based on whether the book was included on a New York Times "recommended" list in the weeks after its review appeared. Each week, the New York Times recommends about eight to twelve – roughly 40 percent – of reviewed books. We also obtain the dates of these recommendations.

For each of the 7,947 titles in the sample, we obtain a list of the books' ISBNs by searching for the title and author on Bowker. We use the ISBNs to retrieve daily Amazon data on the respective editions' sales ranks, prices, number of ratings, and Amazon stars from keepa.com, which provides daily Amazon data, separately for each physical book edition.

We obtain daily Amazon data for the US site as well as two other domains selling English-language books, the Canadian and UK sites. The benefits of these data are considerable for causal identification of rating and review effects. Because we have high-frequency data, we look for high-frequency variation in the sales rank with the appearance of reviews. Moreover, we make use of changes in prices and crowd ratings, all of which can differ across domains as well as over time, to ascertain their impacts on sales.[12] Because we have data on the same edition at different national Amazon domains, we also identify impacts of Amazon

---

[11]We include all reviewers on Goodreads' "most-popular reviewers" lists as of June 2019 who have more than 10,000 followers.

[12]This is another advantage of our data over the Nielsen database. While Nielsen includes list prices, it does not provide information on the prices actually charged for books.

star ratings and prices using cross-platform variation in the changes in, say, ratings and the changes in sales ranks.

Along with these advantages come some disadvantages. First, our data cover only one retailer – Amazon – and not the entire market. Still, Amazon accounted for 44.5 percent of US sales of physical books in 2017 – the year before our sample – so our data cover a major part of the market.[13] Second, we observe the sales rank and not the sales quantity for each edition. We are thus in the position of other authors faced with rank rather than quantity data (e.g. Chevalier and Goolsbee, 2003; Brynjolfsson et al., 2003; Reimers, 2019). Ranks are valuable measures of quantity, but many of our analyses below require a way to translate ranks directly into sales quantities. Amazon does not disclose how it calculates its sales ranks, but a few things are clear.[14] First, many ranks are updated at least daily, often hourly. Second the ranks are not based only on the most recent day. Figure A.3 shows the time series of the Amazon sales rank for a book with modest sales. When a sale occurs, the rank improves sharply, then drifts up for days until the next sale occurs. This clearly indicates that the sales rank is based on a moving average of sales that has a long – multi-day – memory. This will be relevant to both their modeling and their interpretation.

3.2   Summary Statistics

Table 1 provides a description of the sample. The overall sample, in column (1), includes 10,549 distinct editions and just over 1.6 million daily observations across all domains. Columns (2)-(4) report statistics separately for the US, Canada, and the UK. The US sample includes 8,631 editions, and the Canadian and UK samples include about 3,800 editions each. The US sample includes substantially more underlying Amazing ratings.

The inter-quartile range of sample star ratings runs between 4.1 and 4.7. The number of individual ratings underlying these star ratings varies across books and time. By construc-

---

[13]See   https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/78929-print-unit-sales-increased-1-3-in-2018.html.

[14]See https://www.amazon.com/gp/help/customer/display.html?nodeId=525376.

tion, books enter the platform with no ratings, so many titles have star ratings based on few underlying ratings for a time. In the US, the median number of underlying ratings is 60. A quarter of the observations have star ratings based on 15 or fewer, and a tenth of the observations have star ratings based on five or fewer ratings.

We also obtain genre information on sample titles from Bowker. As Figure A.4 shows, the professionally reviewed books have a different genre distribution than the overall sample. The professionally reviewed subsample has higher proportions in "serious" genres such as biography, history, and social science and lower shares in genres such as self-help, romance, and juvenile fiction. Even without deeper analysis, it is clear that the appearance of crowd reviews raises the amount of pre-purchase information available for the genres not attracting the attention of professional review outlets.

3.3   Supplementary Weekly Nielsen Data

We also make use of weekly US sales data from Nielsen Bookscan for three ancillary analyses. We use quantities for the top 100-selling physical editions of each week in 2018 for estimating the elasticity of sales quantities with respect to ranks. We later employ these data for 2015-2018 for estimating a nested logit substitution parameter, which we use in our welfare calculations.

We use a different extract from the Nielsen data for a third exercise, estimating the impact of a consistent subset of New York Times reviews – for the 100 New York Times annual Notable Books – on sales over time, for each even-numbered year between 2004 and 2018.

4   Empirical Strategies and Descriptive Results

We have three goals in this section. First, we provide causal evidence on the relationships between pre-purchase information (reviews and crowd ratings) and sales ranks. Second, we

provide evidence on the potential interaction between the effects of reviews and ratings on particular titles' sales. Third, we translate the measured sales rank coefficients into effects on quantities – such as the elasticity of the quantity sold with respect to the Amazon star rating – that we use to calibrate structural models for welfare analysis.

To accomplish the first goal, we run regressions of log sales ranks on three groups of variables, as well as various fixed effects. The three groups of variables are: a) a lagged value of the log sales rank for the title at the platform; b) indicators for whether a title has recently received a review from a professional outlet; and c) platform-specific measures of Amazon crowd ratings, prices, and the numbers of underlying ratings. To allow for the possibility that star effects vary with the number of underlying ratings, we also include their interaction. Finally, we include country-specific title fixed effects as well as fixed effects for the time until and since the book's publication.

The specifications can be described via the following equation:

$$
\ln(r_{jct}) = \theta \ln(r_{jc,t-1}) + h_{\tau c} + a \ln(p_{jct}) + g \ln(R_{jct}) + m \ln(ratings_{jct}) \tag{1}
$$
$$
+ n \ln(ratings_{jct}) \ln(R_{jct}) + \pi_{\tau'c} + \mu_{jc} + \epsilon_{jct}
$$

In this model, r_{jct}, p_{jct}, R_{jct}, and $ratings_{jct}$ are the sales rank, price, star rating, and number of underlying ratings for title j on platform c on day t. The term $h_{\tau c}$ is a platform-specific coefficient for τ days relative to the appearance of a professional review. Initially, we allow for separate h terms for each of the days leading up to and following the appearance of professional reviews. Finally, the terms $\pi_{\tau'c}$ are fixed effects for the τ' days relative to the book's publication to account flexibly for time patterns of sales around the publication date, and the terms μ_{jc} are platform-specific edition fixed effects.

4.1 Effect Estimates

We first focus on the effect of professional reviews. To that effect, we include the h terms for each of the 20 days before and 100 days after the appearance of a review with the last pre-review day as a baseline.[15] We include two sets of these terms, one for the New York Times and another for the other professional review outlets collectively. We estimate this model on only US data. Figure 2 reports the time patterns of New York Times and other professional review effects. As the left panel shows, a New York Times review delivers a large and immediate improvement in the sales rank when the review appears. The log rank improves by -0.4, then returns to its baseline trend a few weeks later. As the right panel shows, professional reviews at other outlets also have detectable effects, but they are much smaller. The coefficients on the remaining variables are shown in column 1 of Table 2.

Some books reviewed by the New York Times are recommended, and Figure 3 compares the coefficient estimates for recommended vs other books, summarizing the effects with three indicators, for 0-5 days, 6-10 days, and 11-20 days after a review. We also include an indicator that is one from ten days before until 20 days after the appearance of a review so that the post-review effects are defined relative to the ten days before.[16]

The higher sales effect of reviews for books that are later recommended indicates that more positive reviews have larger effects on sales. If we had a continuous measure of review positivity, it would be of interest to determine whether more positive reviews had larger effects on sales and, related, whether the least positive reviews had smaller or even negative effects. To explore this possibility, we used data on the book and the circumstances surrounding the reviews in a LASSO logistic regression to predict the probability that a

---

[15]These effects are identified because reviews do not necessarily appear on the books' publication dates. In our sample, 89.4 percent of New York Times reviews appear more than five days after publication, and 86.4 percent appear more than ten days after. Moreover, effect estimates do not differ between the books reviewed at and after publication.

[16]Some part of the larger effect for recommended books may operate through the appearance of the recommendation. However, the effect for recommended books is larger than for non-recommended books, even when we drop observations after the recommendation appears.

New York Times-reviewed book would also be recommended.[17] While there are some differences between the effects for books with different probabilities of being recommended, these differences are small. Hence, we treat the effect of the New York Times as binary.

We next measure the effects of star ratings, and we continue to represent professional reviews by aggregating days after a review as described above. Estimating the effect of Amazon star ratings is more challenging than estimating review effects, as these ratings evolve less discontinuously, and potentially endogenously, over time. Still, the context gives us promising avenues of identification.

A simple approach to measuring the impact of ratings and prices on sales would be to estimate the cross-sectional relationship between an edition's sales rank and its rating, but the obvious shortcoming of this approach is that books that are "worse" may have both lower ratings and higher (worse) sales ranks, entirely apart from the possible causal impact of ratings on sales. A possible solution to this problem would be to control for the book's unobserved quality, following editions on a particular platform over time. Then the effects of the log price (p) and log star rating (R) on the log sales rank would be identified from the within-title changes.

We implement a variant of this approach in column (2) of Table 2, using only US data. The specification includes both edition fixed effects and the π terms (for time until and since publication). The coefficient on the log star rating is -0.107. This specification also shows impacts of professional reviews that are consistent with those in Figure 2. The appearance of a New York Times review improves the log sales rank by 0.24 in the five days after the

---

[17]In particular, we employed indicators for the day of the week on which the review appeared, along with 85 genre dummies and 372 publisher dummies in a LASSO logistic regression with 10-fold cross validation. The procedure selects 147 of 463 underlying variables and produces a wide range of predicted probabilities that a review will be positive. The inter-quartile range runs from 0.31 to 0.56, and the model accurately predicts which reviews are recommended 72.6 percent of the time. We then interact dummies for quartiles of expected review positivity with the indicators for the first five days after the review, 6-10 days, and 11-20 days. The results are mildly supportive of the approach. While the four estimates for the 0-5 day window are statistically indistinguishable, the effects are slightly larger for the reviews predicted to be more positive: The four effects for the 6-10 day window are, ordered from lowest to highest quartile, -0.08 (se=0.015), -0.14 (0.017), -0.16 (0.019), and -0.14 (0.019). For the 11-20 day window, the corresponding estimates are -0.075 (0.013), -0.069 (0.012), -0.103 (0.014), and -0.108 (0.012).

review; and impacts of other professional reviews are much smaller.

We explore whether the impact of stars depends on the number of underlying ratings by adding an interaction of $\ln(R_{jct})$ and $\ln(ratings_{jct})$, in column (3). The inclusion of the interaction terms shrinks the main star rating effect, and the interaction term itself is negative and significant, indicating that the star ratings have larger effects when they are based on more underlying ratings.

The US-data-only approach of columns (2) and (3) is vulnerable to a concern that some unobserved factor is changing both attitudes toward a title and its sales over time. An alternative is to make use of the differences in ratings changes across platforms to ask whether the cross-platform rating change differential gives rise to a cross-platform sales rank change differential for the same book. This is analogous to an approach that Chevalier and Mayzlin (2006) employ with two time observations. Our data allow us to implement this approach with hundreds of daily observations per title.

Column (4) implements the estimation with all three platforms and platform-specific edition fixed effects as well as the π terms for time until and since publication. Column (5) adds the interaction of $\ln(R_{jct})$ and $\ln(ratings_{jct})$. In these specifications, we interact the h terms for time since a professional review with the platform. This allows the effect of, say, the New York Times to differ across countries. The US New York Times effect for the first five days is about -0.25 in both specifications. While the coefficient on $\ln(R_{jct})$ is smaller in absolute value in column (4) than in column (2), the interaction specifications in columns (3) and (5) give very similar results.

While the interaction specifications appear to fit, we explore whether they are overly restrictive. We replace the $\ln(ratings_{jct})$ term in column (5) with dummies for deciles of the number of underlying ratings and the interactions of these indicators with $\ln(R_{jct})$. Figure 4 reports the results. The magnitude of the coefficient on star ratings rises essentially monotonically in the log number of ratings across rating deciles. While stars have essentially no effect on the sales of books with numbers of ratings in the lowest decile, the coefficient

17

grows in magnitude to -0.50 for the top deciles. We conclude that our interaction specifications provide faithful representations of the underlying relationships. In Appendix section B, we also explore the sensitivity of sales ranks to the visual depiction of stars in half-star increments, as in Luca (2016).

4.2   Star Effects before and after Professional Review Appearance

We use a variant of equation (1) to explore the possible substitution between ratings and reviews. Using all three domains, we include only books that are professionally reviewed, and we interact a post-professional review indicator with the main variables in equation 1, the price, stars, number of ratings, and their interaction. The price coefficient is 0.24 (se=0.010) prior to the review and falls by 0.060 (0.01) afterwards, indicating that demand becomes less elastic. The effect of stars operates through both a direct effect and the interaction with the log number of reviews. The coefficient on log star rating rises from 0.012 (0.024) by 0.047 (0.018), while the coefficient on the interaction falls from -0.082 (0.015) by an additional 0.041 (0.014). Thus, for books with more than 3.1 reviews, the star effect is larger following the appearance of a professional review, reflecting complementarity between reviews and ratings.[18]

4.3   Translating Ranks into Quantities

The evidence above indicates that reviews have an impact on sales ranks, but two steps are required to translate the coefficients into elasticities. First, we need to translate ranks into quantities. While Amazon does not disclose sales quantities, we do have information on the market-wide sales of the top-100 weekly physical bestsellers according to Nielsen. Assuming that sales follow a power law in ranks, i.e. $q_j = A r_j^{-B} e^{\epsilon_j}$ (see Chevalier and Goolsbee, 2003), we can summarize these data by a regression of log quantities on log ranks. This regression, on data for 2018, yields $B = 0.54$ (se=0.004, $R^2 = 0.75$).

---

[18]That is, $3.1 \approx e^{0.047/0.041}$.

Second, equation (1) is a partial adjustment model. We find the full effect of a right hand side variable by setting $\ln(r_{jt}) = \ln(r_{j,t-1})$. Then the derivative of a book's log rank with respect to, say, the log price, from equation (1), is $\frac{a}{1-\theta}$. Combining the above, the reduced form elasticities of quantity with respect to price and the star rating are, respectively:

$$
\begin{aligned}
\epsilon_p &= \frac{\partial \ln(q_j)}{\partial \ln(p_j)} = \frac{aB}{1-\theta}, \\
\epsilon_{Rj} &= \frac{\partial \ln(q_j)}{\partial \ln(R_j)} = \frac{(g + n \ln(ratings_{jt}))B}{1-\theta}.
\end{aligned}
$$

Analogously, the effect of a review on the log sales quantity is $\frac{hB}{1-\theta}$. We obtain standard errors for these estimates by taking 500 parametric bootstrap draws from the estimated joint distributions of the parameters from Table 2, as well as from the distribution of the estimated B.

Table 3 reports estimates of quantity effects from model (5) in Table 2. Rows 2-5 report elasticities of the quantity sold with respect to the Amazon star rating. Because of the interaction of the star rating with the number of underlying ratings, the effect varies with the number of ratings. At the 25th percentile, the elasticity is 0.511, while it is 0.769 at the median and 1.055 at the 75th percentile. At the mean, the elasticity is 0.783. The next rows report the effects of reviews, during particular time windows after their appearance, on log sales. For example, the 0.58 in the NYT first five row indicates that sales increase by 78 percent during the first five days after the appearance of a New York Times review $(e^{0.579} - 1 = 0.78)$.

The bottom panel of Table 3 reports percentage impacts of reviews on annual simulated sales. We estimate daily sales quantities for an edition j by assuming that q_{jt} is proportional to $\frac{1}{exp(ln(rank_{jt})^B)}$ using $B = 0.54$ as described above. We estimate the counterfactual sales absent professional reviews by substituting the following for the log rank:

$$
\ln(rank_{jt}) - \frac{h_k B}{1-\theta}(\text{review indicator } k)_{jt}
$$

19

Here, the indicator k refers to, say, the first five days after the receipt of a New York Times review. We aggregate these estimated quantities across all days in the year, then compare the baseline to the calculated values corresponding to the absence of the respective sources of pre-purchase information. This allows us to calculate the percentage impacts on sales. For example, receiving a New York Times review (but not another professional review) raises sales by 4.63 percent during 2018.

Finally, Table 3 also reports a price elasticity of demand of -0.43. This book-level elasticity appears rather inelastic. We offer four comments at this point. First, the fixed effects in our estimation approaches deal with the endogeneity of prices much as they do for ratings. Second, it is widely understood that Amazon prices below the static profit-maximizing level. In a 2013 60 Minutes interview, Amazon CEO Jeff Bezos stated, "We do price elasticity studies, and every time the math tells us to raise prices."[19] Third, Reimers and Waldfogel (2017) obtain similarly inelastic estimates. Finally, as we will discuss further below, while the absolute size of the welfare effects of pre-purchase information depends on the price coefficient, the relative size of the welfare effects of professional vs crowd reviews – the main study finding – is invariant to it.

Before turning to an explicitly structural approach to welfare analysis, we can get a rough estimate of the relative welfare consequences of reviews and ratings from effects of pre-purchase information on consumers' tendencies to purchase. Weighting professional reviews by their relative frequencies, the average effect of a professional review is to increase purchase by 4.3 percent over the course of the year. With star ratings, the effect depends on the rating. The average title's star rating is 0.08 log points from the cross-title average star rating, which is a consumer's uninformed guess about quality absent pre-purchase information $(\hat{R}_j)$. Given the estimated relationship between star ratings and sales – 0.783 at the mean in Table 3 – stars move the sales of a title by on average 6.3 percent, up or down. On a per-book basis, then, the quantity effects of professional reviews are 47 percent higher. As we detail

---

[19]See https://www.cbsnews.com/news/amazons-jeff-bezos-looks-to-the-future/.

below, the books receiving reviews account for 36 million copies sold in our data, while those receiving star ratings account for 309 million units. Hence, a simple calculation shows that the aggregate effect of star ratings exceeds the impact of reviews by a factor of about 13.[20] This suggests that the aggregate impact of star ratings is substantially higher than the aggregate impact of professional reviews.

5   Welfare Analysis

One of the shortcomings of the foregoing analysis is that despite suggestive calculations, there is no clear way to compare the welfare benefit from the availability of professional reviews versus the existence of Amazon stars. As our theoretical model in section 2 suggests, however, the counterfactual change in consumer surplus provides a natural basis for comparison, and a structural demand model facilitates this calculation. We calibrate our descriptive estimates to a nested logit model of demand. We then estimate the welfare impacts of star ratings and professional reviews, using only data on the US market.

5.1   Preliminaries

In order to build a nested logit model of demand, we need a few components in addition to the descriptive quantity effects estimated above. These include the market size (M), the total 2018 US physical book sales of 695 million, the number of unit sales accounted for by our sample books (192 million), and a nested logit substitution parameter σ, which we estimate to be 0.373. We discuss their derivations in Appendix section C. We also discuss the sensitivity of our results to our σ estimate in the robustness section.

In addition, we need ways of modeling book sales in the absence of reviews and ratings. Our methods for counterfactually simulating the absence of reviews and ratings are different. We assume that a professional review affects consumers' perceptions of the book; and the

---

[20]We obtain 12.6 as (0.063×309/0.043×36).

review effect estimated above reflects the revelation that the book is better than expected. We assume that the sales of reviewed books prior to their reviews indicate how those books would sell in the absence of reviews generally.

Modeling an environment without star ratings is more complicated. Star ratings exert a continuous effect on sales. We assume that in the absence of star ratings, consumers would have some pre-purchase quality perception for each book. This requires a measure of ex ante book quality, denominated in stars, that consumers would have expected absent the stars' existence. We model consumers' beliefs about book quality $(\hat{R}_j)$ via a regression of Amazon stars on publisher fixed effects, genre fixed effects, and dummies for authors' prior experience. For each edition, we use the average star rating across all days. The resulting regression explains 26.25 percent of the variation in stars. We then treat the fitted value as a measure of the ex ante quality of each book that consumers would have expected absent the star rating system. We explore the sensitivity of our results to the explained share of the variation in star ratings in the robustness section below.

Rather than covering all retailers and all books, our data cover one retailer, albeit a major one, and over 10,000 editions. We would like to make statements about market-wide welfare effects of ratings and reviews, and this requires some behavioral and scaling assumptions. First, we assume that effects of professional reviews operate at all retailers, not just at Amazon. Second, we assume that effects of Amazon stars operate only via Amazon sales. In the demand model, we scale the quantities of each book sold to equate the total of sample titles to their market-wide sales. This way, because all professionally reviewed titles are in the sample, professional review results are market totals. To get Amazon star effects, we take our estimate of the change in CS from Amazon stars per dollar of model Amazon revenue. We then multiply this per-dollar measure by Amazon's revenue from sales of all titles (695 million unit sales industry-wide $\times$ \$17.54 per title $\times$ Amazon's 44.5 percent market share).

5.2 A Simple Structural Model

To perform our welfare analysis, we calibrate a nested logit model to the estimated elasticities. We begin by defining

$$\delta_j = \ln(s_j) - \sigma \ln(s_{j|g}) - \ln(s_0), \tag{2}$$

where $s_j = q_j/M$, $s_{j|g} = q_j/Q$, and $s_0 = 1 - Q/M$. Note that $Q = \sum q_j = 192$ million units sold market-wide for the sample titles, as discussed in Appendix section C. Each product's share is then

$$s_j = \frac{e^{\delta_j/(1-\sigma)}}{1 + \sum e^{\delta_{j'}/(1-\sigma)}} \frac{D^{1-\sigma}}{1 + D^{1-\sigma}},$$

where $D = \sum e^{\delta_j/(1-\sigma)}$ (see Berry, 1994).

Let δ_j be the utility in the status quo, when reviews and ratings are present. We can write this as

$$\delta_j = \delta_j^0 + \alpha p_j + \gamma_j R_j + \psi_j,$$

where α, γ_j, and ψ_j are utility function parameters. While these parameters are unknown, they are related to estimated parameters a, g, h, and n from equation (1), respectively. We can calculate the nested logit expressions for the derivatives of quantity with respect to price and rating, set these equal to the reduced form derivatives described above, then solve for the utility function parameters.

In particular, the nested logit model gives a simple expression for the price elasticity of demand:

$$\hat{\epsilon}_p = \alpha_j \frac{p_j}{1 - \sigma} \left(1 - \sigma s_{j|g} - (1 - \sigma)s_j\right)$$

Given $\hat{\epsilon}_p$ from the descriptive analysis above, σ (from the appendix), and $s_{j|g}$, s_j, and p_j (which are data), this formula gives us a parameter estimate of α_j for each j, which we average to obtain our estimate of utility function parameter α. Analogously, we can infer γ_j

by solving the following equation:

$$\hat{\epsilon}_{Rj} = \gamma_j \frac{R_j}{1-\sigma} \left(1 - \sigma s_{j|g} - (1-\sigma)s_j\right)$$

Rather than taking a simple average of the γ_js, we accomodate the interaction of stars and the numbers of ratings by solving the structural model using a γ_j that is parameterized as a function of the number of reviews that each book has received.

We can solve for the utility function parameters associated with reviews in a related way. Our descriptive analysis tells us how each sales quantity q_j would have been different in the absence of reviews, q'_j. The model analogue of our descriptive measure $\ln(q_j/q'_j)$ is the review-induced percentage change in sales for reviewed books, relative to the review-induced percentage change in sales for unreviewed books. Define s^r_j as sales of reviewed books in the presence of reviews, s^u_j as sales of unreviewed books in the presence of reviews, s'^r_j as sales of reviewed books in the absence of reviews, and s'^u_j as sales of unreviewed books in the absence of reviews. Then the descriptive fact is related to its model analogue via

$$\ln(q_j/q'_j) = \ln(s^r_j/s'^r_j) - \ln(s^u_k/s'^u_k),$$

where j indexes reviewed books, and k indexes non-reviewed books. A few lines of algebra show that in our nested logit model, $\ln(q_j/q'_j) = \frac{\psi_j}{1-\sigma}$. Given σ, we therefore know ψ_j; and we parameterize ψ_j by dividing books into four groups, those reviewed by the New York Times, those reviewed by other professional outlets, those reviewed by both, and a baseline group reviewed by neither.

Given values of the utility function parameters, we can compare two counterfactual scenarios – without Amazon star ratings and without professional reviews – to the baseline when both are present.[21] We are interested in the effects of the two sorts of pre-purchase

---

[21]We leave prices unchanged in these counterfactual scenarios because unreported estimation exercises show negligible effects of either type of pre-purchase information on prices.

information on the consumer surplus achieved in the market. First, we need expressions for the status quo utility level, as well as its analogues in the absence of crowd and professional reviews. Status quo utility of product j is given by the data, as in equation (2), while counterfactual utility absent Amazon stars is given by $\delta_j^s = \delta_j - \gamma_j(R_j - \hat{R}_j)$; and counterfactual utility absent professional reviews is given by $\delta_j^p = \delta_j - \psi_j$.

The change in CS associated with star ratings is given by

$$\Delta CS = \frac{M}{\alpha}\left[\ln\left(1 + \sum \exp\left(\frac{\delta_j}{1-\sigma}\right)^{1-\sigma}\right) - \ln\left(1 + \sum \exp\left(\frac{\delta_j^s}{1-\sigma}\right)^{1-\sigma}\right) - \sum \gamma_j(R_j - \hat{R}_j)s_j^s\right],$$

where the term $\sum \frac{\gamma_j}{\alpha}(R_j - \hat{R}_j)s_j^s$ is an adjustment reflecting the possibility that what is consumed has ex post utility that differs from the ex ante value (such as area B in Figure 1), and s_j^s is the market share of product j in the absence of star ratings (where $M s_j^s$ corresponds to Q_1 in Figure 1).

The respective change in consumer surplus from the presence of professional reviews is given by the analogous equation, with δ_j^s replaced by δ_j^p, and with $\sum \gamma_j(R_j - \hat{R}_j)s_j^s$ replaced by $\sum \psi_j s_j^p$, where s_j^p is the market share of product j in the absence of professional reviews.

5.3   Results

Table 4 shows the revenue and consumer surplus effects. Both forms of pre-purchase information have impacts on sales and revenue. Professional reviews raise revenue by $28.12 million. On net, the presence of star ratings raises book revenue by $52.47 million across all books – about twice as much as do professional reviews. For the "unexpectedly good" titles with star ratings that beat expectations, pre-purchase information raises sales by $191.11 million, while it decreases sales of "unexpectedly bad" titles by $138.64 million.[22]

Professional reviews, by shifting consumption to "good" titles, raise consumer surplus by $2.88 million. Compared with the counterfactual environment without star ratings, the

---

[22]The absolute value of the star rating-induced revenue change exceeds the professional review-induced change by a factor of 11.7, which is similar to our suggestive welfare calculation in section 4.

status quo delivers $41.25 million in additional consumer surplus from Amazon sales alone, which is roughly 15 times the effect of professional reviews on CS through all retail channels. Thus – and this is our main result – the vast majority of the value of pre-purchase information available to consumers following digitization is delivered by crowd-based rating systems.

Some part of this result is driven by the wider availability of star ratings than professional reviews. It is of interest to compare the relative effect sizes on a comparable set of titles. To this end, we perform a welfare calculation based on counterfactual removals of the star ratings on the professionally reviewed books (leaving the star ratings in place on the others). As Table 4 shows, the presence of star ratings on the reviewed books adds $1.80 million, in comparison to the $2.88 million added by the reviews. On a per-book basis, then, the dollar value of the impact of professional reviews is about 60 percent higher.

5.4 Robustness

Our welfare analyses depend on the estimated parameters. Here we explore the sensitivity of our basic results to different values of the parameters α, B, and σ, to the consumers' ability to predict a book's true star rating, and to allowing rating and price effects to change after reviews.

The price parameter α determines the absolute size of a welfare effect. It does not, however, affect the relative sizes of the respective effects of professional reviews and crowd ratings on consumer surplus, so our conclusions on the relative impacts of professional reviews and star ratings are unaffected by α. The same logic applies to the power law parameter B.

The impact of the substitution parameter σ on ΔCS is less obvious a priori. To the extent that counterfactual changes resemble increases in market size, we would expect smaller impacts on welfare, the greater the substitutability of products. In our star rating counterfactual, some products' mean utilities rise while others fall, so that the effective market size does not change substantially. Hence, our results change only minimally across a range of σ values between 0 and 0.9.

The measured welfare benefits of Amazon star ratings also depend on the accuracy of consumers' predictions of product quality absent star ratings. Our baseline model of this is a regression of stars on observables, and the regression explains 26.25 percent of the variation. It is possible that the regression understates, or overstates, the ability of consumers to predict quality. We explore the sensitivity of our result to prediction accuracy using the approach of Aguiar and Waldfogel (2018). We add the following explanatory variable to the star rating regression: $R_j + \kappa \cdot \epsilon_j$, where ϵ_j is a standard normal random error, and κ is a scale factor we vary to produce variation in the prediction accuracy, which we summarize by the R^2 of the regression. Figure 5 shows how the change in CS from the presence of Amazon stars varies with prediction accuracy, with dots for our baseline estimate and for the zero-information case ($\hat{R} = \bar{R}$, for $R^2 = 0$). If our model understates prediction accuracy, then the true welfare benefit is lower. For example, if prediction accuracy corresponded to an R^2 of 50 percent, then the welfare benefit would be roughly \$20 million. If consumers could perfectly predict quality absent star ratings ($R^2 = 100$ percent), star ratings would deliver no welfare benefit. Consumers would need to be very well informed in the absence of reviews and ratings – an R^2 of 80 percent – in order for Amazon stars to add as little consumer benefit as professional reviews.

Finally, we also perform the welfare analysis above using quantity measures derived from a specification allowing for different price and star rating effects following the appearance of professional reviews. The resulting changes in CS for stars and for professional reviews are \$35.0 and \$1.69 million, respectively, for a ratio of roughly 20 to 1.

6 Professional Review Effects over Time

We saw earlier that star effects rise after titles have received professional reviews. This suggests that, at the title level, ratings and reviews are complementary; but the growing availability of star ratings may have changed the effect of reviews. To explore this, we would

ideally repeat the foregoing analyses for earlier years, prior to the diffusion of online retail and associated crowd ratings. This is infeasible, however, because our daily ranking data do not reach back far enough. Still, we can employ the approach of Berger et al. (2010) and use weekly physical book sales data to estimate the impact of New York Times book reviews on demand.

For this analysis, we assemble data on a comparable list of books over time, the 100 Notable Books for even-numbered years from 2004 to 2018. We manually search for these books' ISBNs in the Nielsen Bookscan database to collect weekly unit sales, and we obtain their review dates from the New York Times.[23] We then estimate regressions of the form:

$$\ln\left(\frac{s_{jt}}{s_{j,t-1}}\right) = \lambda \text{review}_{jt} + \beta x_{jt} + u_{jt},$$

where s_{jt} denotes the sales of book j in week t, $\text{review}_{jt} = 1$ in the week immediately following the New York Times review and x_{jt} includes controls for the number of weeks since the book's release. We also include a dummy variable that equals one in all weeks after publication. Like Berger et al. (2010), we drop all observations more than nine weeks before or after the review. The form of the dependent variable means that our coefficient of interest, λ, measures the impact of a review on the rate of change of sales.

Figure 6 presents coefficient estimates. We find that professional reviews had positive and significant effects on sales in all years. The size is roughly constant through 2016, then more than doubles in 2018. If anything, the effect of the New York Times book reviews has increased over the last 15 years. This may be due to a different aspect of digitization: The number of digital-only subscriptions to the New York Times rose from about 100,000 in March 2011 to over 2.5 million in the third quarter of 2018 (Richter, 2018).

---

[23]We limit our analysis to the list of notable books because their reviews are likely most positive, and because manually searching for ISBNs is quite time consuming. We obtain these books' ISBNs from Goodreads.

7 Conclusion

Digitization has delivered a large number of new products, straining the capacity of both critics and consumers to discover those meriting their attention. At the same time, digitization has delivered a potential solution in new mechanisms for aggregating user product ratings into potentially useful pre-purchase information for other consumers. Using Amazon daily data on sales ranks, prices, and star ratings for over 10,000 book editions, along with information on review timing in professional review outlets, we document causal impacts of pre-purchase information on sales ranks. We then transform these estimates into impacts on quantities, which we use to calibrate nested logit demand models for welfare analysis. We find that book reviews in the New York Times and other major newspapers have substantial impacts on book sales. New York Times reviews raise sales by over 75 percent in the five days after a review and by 4.6 percent over the year. We also document that the elasticity of quantity sold with respect to Amazon stars averages about 0.75.

Because these two forms of pre-purchase information have causal impacts on buying behavior, they also affect welfare. Professional reviews raise revenue by $28.12 million and raise consumer surplus by $2.88 million. Crowd ratings have net impacts on revenue that are twice as large as the net revenue impacts of professional reviews, but the difference is even larger for consumer surplus. The existence of crowd ratings adds $41.25 million to consumer surplus from Amazon book purchases alone, or 15 times the impact of professional reviews on surplus derived from purchases through all channels. We conclude that digitization, in addition to delivering a proliferation of new products, has also added substantially to the value of the pre-purchase information available to consumers. Much of the newly available information covers books in genres overlooked by professionals. Still, while the value of crowd-based pre-purchase information to consumers is now much larger than the consumer benefit derived from professional reviews, the absolute impacts of professional reviews have not declined over time.

Crowd ratings are available for all products and not just for books; particularly given the

smaller role of professional reviewers for other products, crowd-based ratings made available by digitization may also add substantial benefits for consumers of other products. This seems a fruitful topic for further research.

References

AGUIAR, L. AND J. WALDFOGEL (2018): "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music," *Journal of Political Economy*, 126, 492–524.

ALLCOTT, H. (2011): "Consumers' perceptions and misperceptions of energy costs," *American Economic Review*, 101, 98–104.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "High-dimensional methods and inference on structural and treatment effects," *Journal of Economic Perspectives*, 28, 29–50.

BERGER, J., A. T. SORENSEN, AND S. J. RASMUSSEN (2010): "Positive effects of negative publicity: When negative reviews increase sales," *Marketing Science*, 29, 815–827.

BERRY, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

BRYNJOLFSSON, E., Y. HU, AND M. D. SMITH (2003): "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, 49, 1580–1596.

CHEVALIER, J. AND A. GOOLSBEE (2003): "Measuring prices and price competition online: Amazon. com and BarnesandNoble. com," *Quantitative marketing and Economics*, 1, 203–222.

CHEVALIER, J. A. AND D. MAYZLIN (2006): "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, 43, 345–354.

DARGIS, M. (2014): "As indies explode, an appeal for sanity: flooding theaters isn't good for fimmakers or filmgoers," *New York Times*, 9.

DEUTSCHMAN, A. (2004): "The Kingmaker Walt Mossberg makes or breaks products from his pundit perch at a little rag called The Wall Street Journal," *WIRED-SAN FRANCISCO-*, 12, 120–132.

DUAN, W., B. GU, AND A. B. WHINSTON (2008): "Do online reviews matter?—An empirical investigation of panel data," *Decision support systems*, 45, 1007–1016.

FORMAN, C., A. GHOSE, AND B. WIESENFELD (2008): "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information systems research*, 19, 291–313.

GARTHWAITE, C. L. (2014): "Demand spillovers, combative advertising, and celebrity endorsements," *American Economic Journal: Applied Economics*, 6, 76–104.

GREENFIELD, J. (2012): "Seven Advantages Barnes & Noble has in the Bookseller Wars," *Digital Book World*.

HELMERS, C., P. KRISHNAN, AND M. PATNAM (2019): "Attention and saliency on the internet: Evidence from an online recommendation system," *Journal of Economic Behavior & Organization*, 161, 216–242.

JIN, G. Z. AND A. T. SORENSEN (2006): "Information and consumer choice: the value of publicized health plan ratings," *Journal of health economics*, 25, 248–275.

LUCA, M. (2016): "Reviews, reputation, and revenue: The case of Yelp. com," *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*.

MARTIN, A. (2011): "The End of the Career Food Critic," *https://www.theatlantic.com/national/archive/2011/09/sam-sifton-departure-and-the-end-of-the-career-food-critic/337844/*.

NARULA, S. K. (2014): "Millions of people reading alone, together: The rise of Goodreads," *The Atlantic*.

POMPEO, J. (2017): "Michiko Kakutani, the Legendary Book Critic and the Most Feared Woman in Publishing, is Stepping Down from the New York Times," *Vanity Fair*.

REIMERS, I. (2019): "Copyright and generic entry in book publishing," *American Economic Journal: Microeconomics*, 11, 257–84.

REIMERS, I. AND J. WALDFOGEL (2017): "Throwing the books at them: Amazon's puzzling long run pricing strategy," *Southern Economic Journal*, 83, 869–885.

REINSTEIN, D. A. AND C. M. SNYDER (2005): "The influence of expert reviews on consumer demand for experience goods: A case study of movie critics," *The journal of industrial economics*, 53, 27–51.

RICHTER, F. (2018): "The "Failing" NY Times Passes 2.5 Million Digital Subscriptions," `https://www.statista.com/chart/3755/digital-subscribers-of-the-new-york-times/`.

SENECAL, S. AND J. NANTEL (2004): "The influence of online product recommendations on consumers' online choices," *Journal of retailing*, 80, 159–169.

SORENSEN, A. T. (2007): "Bestseller lists and product variety," *The journal of industrial economics*, 55, 715–738.

TRAIN, K. (2015): "Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples," *Journal of choice modelling*, 16, 15–22.

WALDFOGEL, J. (2017): "How digitization has created a golden age of music, movies, books, and television," *Journal of Economic Perspectives*, 31, 195–214.

WALDFOGEL, J. AND I. REIMERS (2015): "Storming the gatekeepers: Digital disintermediation in the market for books," *Information economics and policy*, 31, 47–58.

Figure 1: Illustration – welfare analysis of pre-purchase information
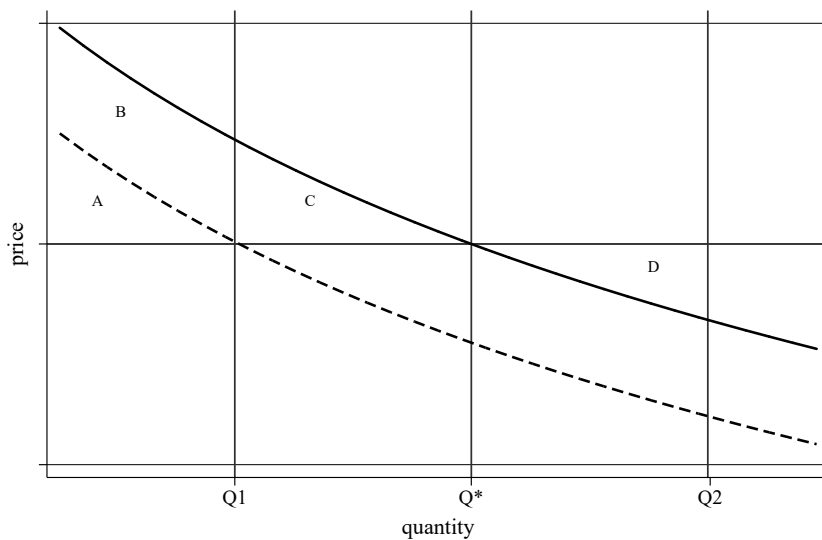
Notes: This figure illustrates demand curves under full information about a product's quality (solid line) and with limited ex ante information about the quality when the expected quality is less than the true quality (dashed line). The corresponding consumer surplus under full information is areas $A + B + C$; under limited ex ante information, it is $A + B$.

Figure 2: Daily effects of professional reviews on sales ranks
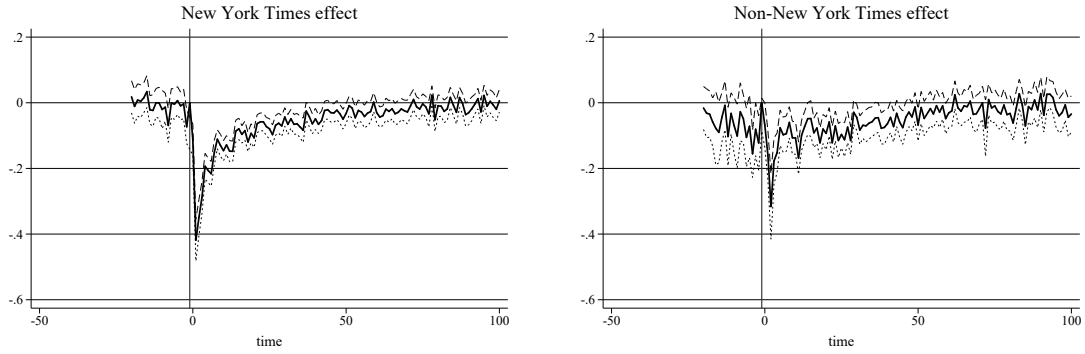
Notes: These figures show coefficients and 95% confidence intervals for each day before and after a New York Times (left panel) and other major review (right panel). The estimates are from a regression of log rank on its lag, price, stars, and the number of underlying ratings, in addition to title fixed effects and days until and since publication dummies, using Amazon US data.

Figure 3: Effect of New York Times reviews – recommended vs. not
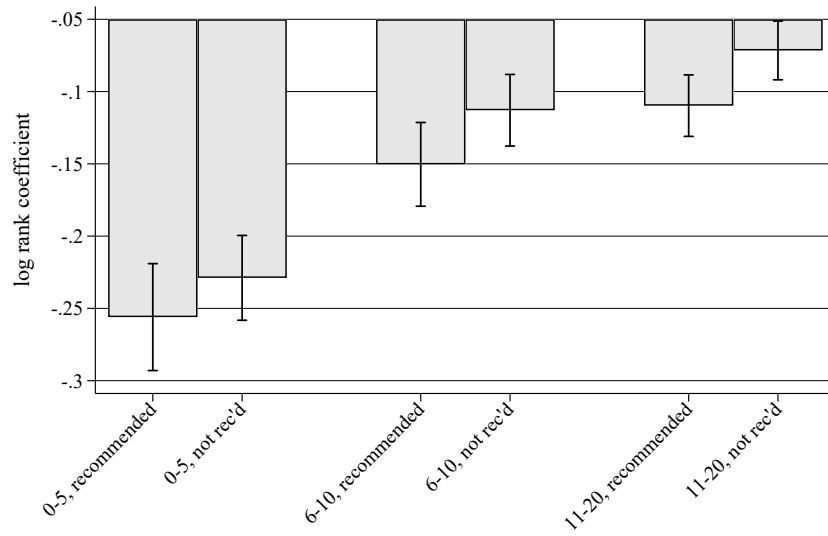
Notes: This figure shows coefficients and 95% confidence intervals for days after a New York Times review, separately for books that were recommended and those that were not. The estimates are from a regression of log rank on its lag, price, stars, and the number of underlying ratings, in addition to title fixed effects and days until and since publication dummies, using Amazon US data.

Figure 4: Effects of Amazon star ratings by ratings decile
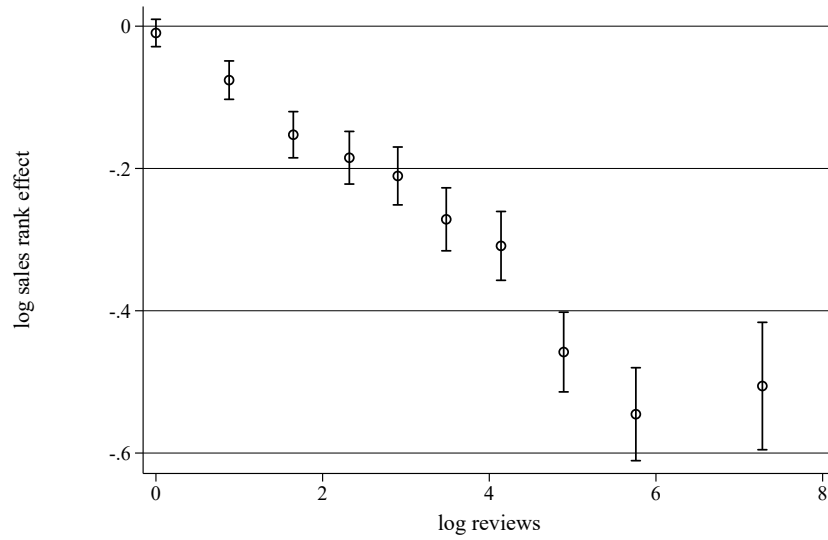
Notes: This figure shows coefficients and 95% confidence intervals for a book's star rating interacted with each decile of the number of underlying ratings, in a regression of log rank on its lag, price, days since various professional reviews, and the number of underlying ratings, in addition to title-platform fixed effects and days until and since publication dummies, using Amazon data from the US, Canada, and Great Britain.

Figure 5: Prediction accuracy and the welfare benefit of crowd information
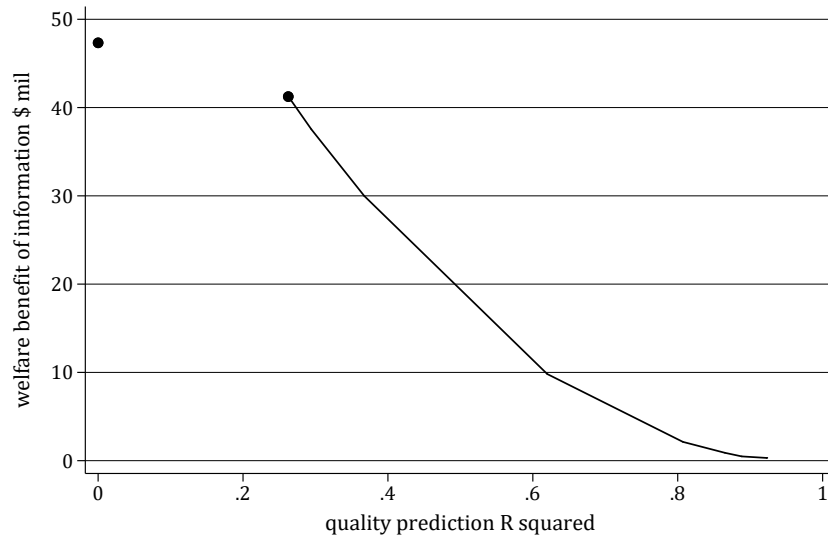
Notes: This figure depicts the estimated welfare gains from the existence of a star rating system for varying levels of explained variation in star ratings, as measured by the R^2 of a regression of the book's true star rating on genre and publisher dummies as well as author experience controls. We increase the R^2 by adding the following explanatory variable to the regression: $R_j + \kappa \cdot \epsilon_j$, where ϵ_j is a standard normal random error and κ is a scale factor of varying size. The leftmost dot ($R^2 = 0$) assigns the overall average star rating to all books. The baseline accuracy, $R^2 = 0.2625$, is denoted by the second dot from the left.

Figure 6: Effects of New York Times reviews from 2004 to 2018
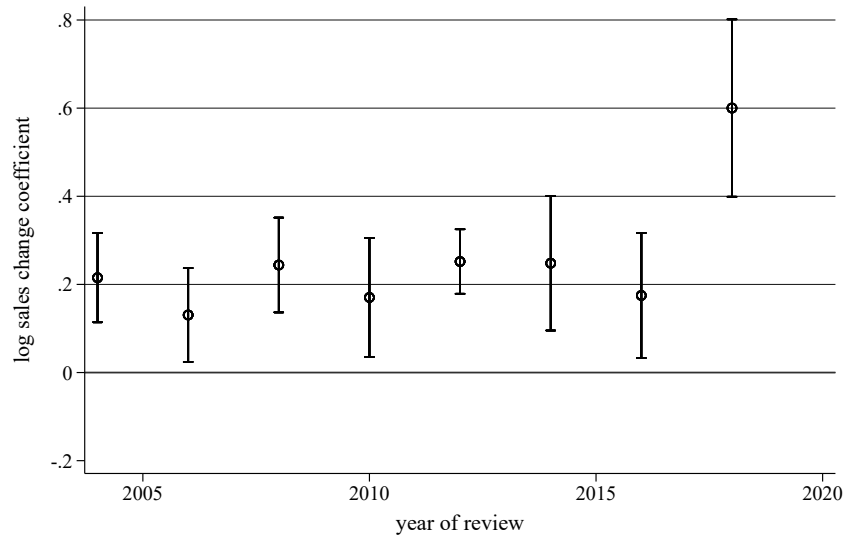
Notes: This figure displays coefficients and 95% confidence intervals for the review dummy in regressions of $\ln\left(\frac{s_{jt}}{s_{j,t-1}}\right)$ on a dummy for the week after a New York Times review and controls for the number of weeks since the book's publication. The regressions were done separately for each even review year from 2004 to 2018, on all books on the New York Times Notable Books lists for their respective years.

Table 1: Sample characteristics

	all	Canada	Great Britain	US
price	18.12	24.73	14.60	17.54
star rating	4.38	4.39	4.35	4.39
sales rank	448,308	203,825	639,051	450,765
reviews	326.25	27.73	111.45	472.00
star rating percentiles				
10th	3.7	3.6	3.6	3.8
25th	4.1	4.1	4	4.2
50th	4.5	4.5	4.5	4.5
75th	4.7	4.9	4.8	4.7
90th	5	5	5	4.9
titles	7,947	3,774	4,032	7,759
editions	10,549	4,340	4,543	9,869
observations	1,612,014	264,452	325,797	1,021,765

Notes: Average prices, star ratings, sales ranks, and number of reviews, across all days in 2018. The samples include all editions of books that entered the 2018 USA Today weekly bestseller lists, that were reviewed professionally in 2018, or that were reviewed by highly-followed Goodreads reviewers. Column (1) includes all books and all platforms (Amazon US, Canada, and Great Britain). Columns (2)-(4) include all books on each individual platform.

Table 2: Effects of crowd and professional reviews on log sales ranks

	(1)	(2)	(3)	(4)	(5)
lagged log sales rank	0.785	0.786	0.786	0.760	0.760
	(0.000766)	(0.000764)	(0.000764)	(0.000656)	(0.000656)
log Amazon price	0.194	0.193	0.195	0.188	0.190
	(0.00406)	(0.00406)	(0.00406)	(0.00307)	(0.00307)
log # of reviews	0.0484	0.0488	0.171	0.0376	0.160
	(0.00125)	(0.00125)	(0.00846)	(0.00100)	(0.00739)
log star rating	-0.108	-0.107	-0.00354	-0.0680	-0.00323
	(0.0118)	(0.0118)	(0.0142)	(0.00860)	(0.00971)
log reviews $\times$ log stars			-0.0829		-0.0827
			(0.00568)		(0.00497)
US: NYT, 0-5 days		-0.240	-0.241	-0.257	-0.257
		(0.0129)	(0.0129)	(0.0130)	(0.0130)
US: NYT, 6-10 days		-0.129	-0.130	-0.152	-0.153
		(0.0110)	(0.0110)	(0.0111)	(0.0111)
US: NYT, 11-20 days		-0.0883	-0.0897	-0.101	-0.103
		(0.00931)	(0.00930)	(0.00938)	(0.00937)
CA: NYT, 0-5 days				-0.112	-0.111
				(0.0422)	(0.0422)
CA: NYT, 6-10 days				-0.0640	-0.0633
				(0.0407)	(0.0407)
CA: NYT, 11-20 days				-0.00996	-0.0107
				(0.0348)	(0.0348)
GB: NYT, 0-5 days				-0.0298	-0.0294
				(0.0259)	(0.0259)
GB: NYT, 6-10 days				-0.000492	-0.000872
				(0.0240)	(0.0240)
GB: NYT, 11-20 days				0.0130	0.0129
				(0.0207)	(0.0207)
US: other, 1-10 days		-0.0300	-0.0294	-0.0370	-0.0364
		(0.0173)	(0.0173)	(0.0175)	(0.0175)
US: other, 11-20 days		0.000674	0.00115	-0.00114	-0.000516
		(0.0165)	(0.0165)	(0.0167)	(0.0167)
CA: other, 1-10 days				0.0414	0.0454
				(0.0706)	(0.0706)
CA: other, 11-20 days				0.0711	0.0759
				(0.0680)	(0.0680)
GB: other, 1-10 days				0.0601	0.0584
				(0.0612)	(0.0612)
GB: other, 11-20 days				0.0742	0.0759
				(0.0616)	(0.0615)
Observations	1,021,765	1,021,765	1,021,765	1,612,014	1,612,014
R-squared	0.968	0.968	0.968	0.961	0.961

Notes: Regression of Amazon log daily sales rank on its one-day lag, as well as the log price, log number of reviews, the log of the star rating, and indicators for whether the title had recently been reviewed by the New York Times or another major US outlet. The first three columns include only data from Amazon's US site. Columns (4) and (5) include data from Amazon's US, Canadian, and Great Britain sites. All specifications include country-specific title fixed effects as well as fixed effects for days until and since publication. Robust standard errors in parentheses.

Table 3: Causal quantity effects

	effect	se
Price elasticity	-0.427	0.007
Amazon stars elasticity (25th pctile)	0.511	0.028
Amazon stars elasticity (50th pctile)	0.769	0.040
Amazon stars elasticity (75th pctile)	1.055	0.056
Amazon stars elasticity (mean)	0.783	0.041
NYT 0-5	0.579	0.029
NYT 6-10	0.345	0.025
NYT 11-20	0.231	0.021
Other 0-10	0.082	0.039
Other 11-20	0.001	0.038
% effect of review on annual q		
Other only	0.566	0.494
NYT only	4.633	0.244
both	5.723	0.694
average	4.310	

Notes: The price and Amazon star rows show estimated elasticities of quantity sold with respect to price and Amazon stars, respectively, based on column (5) of Table 2. The Amazon stars elasticities are reported according to percentiles of the number of underlying ratings. The NYT and Other rows show percentage impacts of reviews on sales during the relevant numbers of days after the reviews at the New York Times and other newspapers, respectively. The bottom panel shows the percentage impacts of being reviewed in the New York Times or other professional outlets on estimated sales over the year. Standard errors are based on 500 parametric bootstrap replications. We draw from the estimated joint distributions of the parameters from Table 2, as well as from the distribution of B from $q_j = Ar_j^{-B}$ using the Nielsen data.

Table 4: Welfare impacts of professional reviews and Amazon star ratings

	Stars	Reviews
ΔRevenue (net)	52.47	28.12
	(5.92)	(2.64)
$R > \hat{R}$	191.11	
	(9.99)	
$R < \hat{R}$	-138.64	
	(7.65)	
ΔCS (total)	41.25	2.88
	(4.20)	(0.35)
ΔCS (reviewed books)	1.80	2.88
	(0.18)	(0.35)

Notes: All dollar figures in millions. Baseline revenue is \$3,057.74 million. The scaled dollar figures for star ratings are calculated by multiplying their impact on CS per dollar spent by our estimate of the spending on physical books at Amazon in 2018 (695 million volumes sold during 2018, times their average price of \$17.54, times Amazon's 44.5 percent share of the market). Because we include all of the books reviewed at the New York Times and the other major papers in the sample, the model's direct measure of the change in CS from these reviews requires no scaling. Figures are based on estimates in column 5 of Table 2. Standard errors are based on 100 parametric bootstrap draws.

Appendix

A   Goodreads Amateur Review Impacts

In addition to making crowd opinion visible, digitization has also fostered growth in opportunities for amateurs to distribute their book reviews. Goodreads is a site where readers rate and review books. It was founded in 2007 and has grown quickly. According to Narula (2014), Goodreads had 25 million users by 2014, and the number of registered users reached 80 million in November 2018. As of April 2016, Goodreads reported that 50 million reviews had been posted to the site. Leaving reviews, and having substantial numbers of followers, are relatively uncommon; but some users have large numbers of followers and leave substantial numbers of reviews. Hence, the appearance of reviews by users with substantial numbers of followers can provide crowd analogues to reviews in critical outlets.

It is difficult to say exactly how visible Goodreads reviews are in comparison with major professional review sources. Based on Similarweb traffic data for December 2018, Goodreads received 98.6 million visits, compared with 302.5 million for the New York Times' entire website.

To explore possible impacts of Goodreads reviews on sales, we assemble a list of reviews from Goodreads' "most popular" reviewer lists and include those with more than 10,000 followers. This produces a list of 1,742 reviews of titles reviewed and published during 2018.

We measure their impacts on sales ranks using the two approaches employed in the paper. First, we measure the daily impacts for 20 days before and 100 days after the appearance of a highly followed reviewer's review. This is analogous to our approaches in Figure 2 in the paper. Appendix Figure A.1 shows the result: There is no apparent effect of these amateur reviewers' reviews on Amazon US sales ranks.

We then measure the impacts for three discrete periods following the appearance of these Goodreads reviews: 0-5 days, 6-10 days, and 11-20 days, also analogous to our approaches in the US-data-only columns of Table 2. We find insignificant effects (coefficients of -0.008 (se=0.008), -0.011 (0.008), and -0.009 (0.006)). We also checked whether the effects were different for the books receiving 5-star ratings, but the coefficients were statistically indistinguishable.

B   Star-Rating Discontinuities

The way that Amazon reports its star ratings gives rise to an additional identification strategy for measuring the impact of star ratings on sales. On a book's page, a customer sees an image of the number of stars that is denominated in half stars, but if one hovers over the star image, one sees a number of stars to a single decimal place. It is easy for a user to see the decimal star rating, but the visual, half-star image may have additional salience. This suggests an additional, discontinuity method for identifying the impact of stars that is reminiscent of Luca (2016). We look for jumps

**Figure A.1:** Goodreads daily effects

Goodreads effect
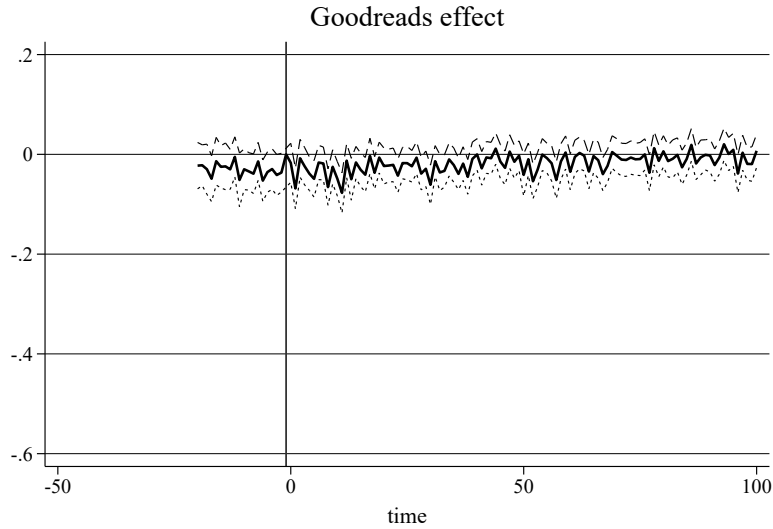
Notes: The figure shows coefficients and 95% confidence intervals for each day before and after a Goodreads top reviewers' review. The estimates are from a regression of log rank on its lag, price, stars, and the number of underlying ratings, in addition to title fixed effects and days until and since publication dummies, using Amazon US data.

in log sales ranks at the decimal star ratings for which the visual half stars jump by one half. This occurs, for example, at 2.8, 3.3 and 3.8 stars. To explore this, we estimate a variant of model (5) above in which we include a series of dummies for each of the possible decimal star ratings instead of the continuous measure $\ln(R_{jt})$. Figure A.2 displays the pattern of coefficients on the decimal rating dummies, with vertical lines at the discontinuities. Ninety percent of these ratings fall between 3.4 and 5, so we focus on this range. We see an overall pattern of better ranks as the rating increases, with larger jumps at the discontinuities. We take this as additional evidence that star ratings have a causal impact.

C   Logit Preliminaries

We obtain the market sizes and collective unit sales of the books in our sample as follows. First, for market size, we assume that each member of the US population is making a monthly decision of whether to purchase a book, so $M = 12 \times 327$ million. Second, we estimate the annual sales per sample title as follows. From data outside our sample, we know the total physical sales for the year

Figure A.2: Effects of Amazon star ratings on sales
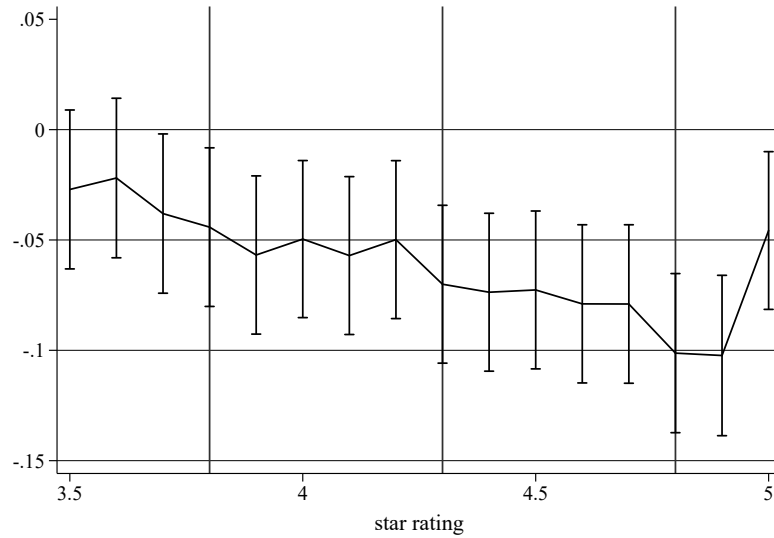
Notes: This figure displays coefficients and 95% confidence intervals for each star rating dummy (from 3.5 to 5 stars) in a regression of log rank on its lag, price, days since various professional reviews, and the number of underlying ratings, in addition to title-platform fixed effects and days until and since publication dummies, using Amazon data from the US, Canada, and Great Britain.

(695 million units).[24] Combined with the (calculated) 2018 unit sales of the Nielsen weekly top 100 books (68.3 million), this implies that the top 100 weekly titles account for 9.84 percent of total physical sales. Because our estimation sample includes only rank data, even with an estimated elasticity of the quantity sold with respect to the sales rank (B), we can only estimate daily sales of each title up to a scalar A, i.e. $q_{jt} = A/\ln(r_{jt}^B)$. But we can choose A so that the sum of weekly top 100 sales in the sample equals the Nielsen total of 68.3 million. This in turn indicates that our sample titles collectively account for 192 million units.

---

[24]See https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/78929-print-unit-sales-increased-1-3-in-2018.html.

Table A.1: Model inputs

variable	value
Market-wide unit sales	
2018 US unit sales (mil)	695
2018 unit sales of Nielsen top 100 (mil)	67.4
Nielsen top 100 share of US sales	0.0984
top 100 share of sample sales	0.3557
US unit sales of sample titles (mil)	192.28
Substitution parameter σ	0.373 (0.0557)
Amazon unit sales	
share of physical sales (2017)	0.455
all titles (mil)	316.2
sample titles (mil)	81.6
Overall market size	
US pop 2018 (mil)	327.2
market size (12*pop)	3,926.4

Notes: This table reports inputs for the nested logit model in section 5. 2018 US unit sales are from `https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/78929-print-unit-sales-increased-1-3-in-2018.html`; 2018 unit sales of Nielsen top 100 are calculated from the Nielsen Bookscan database. US unit sales of the titles in our sample are $Q = \frac{1}{0.3557} \times 0.0984 \times 695$. The substitution parameter σ is estimated using weekly industry demand and unit sales of bestsellers. Amazon's share of physical sales: `https://www.idealog.com/blog/changing-book-business-seems-flowing-downhill-amazon/`. Amazon unit sales of all titles and our sample titles are market-wide sales times Amazon's share.

Estimating the Nested Logit Parameter We infer the degree of substitutability using the Nielsen top 100 sales weekly data, which we have for 2015-2018. For this purpose, we need a few additional pieces of information, along with an instrumental variables strategy. We describe these in turn.

First, we obtain weekly data on total physical book sales from Publishers Weekly, which reports this in most but not all weeks. We refer to this as Q_t. We have these data for 124 weeks during 2015-2018. First, as above, we assume that each member of the US population is making a monthly decision of whether to purchase a book, so with weekly observations, the market size is $M = 0.25 \times 327$ million.

We then define the following variables:

$$s_{jt} = q_{jt}/M, \qquad s_{jt|g} = q_{jt}/Q_t, \qquad s_{0t} = 1 - Q_t/M.$$

As in Berry (1994), we seek to obtain σ from a regression of $\ln(s_j) - \ln(s_0)$ on $\ln(s_{jt|g})$. Intuitively, identification comes from the relationship between the number of products available and whether the share of the population buying books increases.

There is seasonality in the book market, with a substantial increase in sales around Christmas. Publishers know this and may release more books around Christmas, raising a concern that the number of books coming out as well as demand might rise around Christmas. This would look like an effect of product entry on market expansion, even if it were not. To address this, we include week-of-the-year dummies.

Second, we need an instrument for the books' inside shares $s_{jt|g}$. One natural idea would be the number of products available in each week. In our data it is by construction 100. More to the point, however, not all products are of equivalent importance. We appeal to the logic of BLP instruments, which are terms involving the other products in the choice set. Here, for example, we measure the number of products in the top 100 that were originally released in the past week, 2 weeks, and so on, up to ten weeks. Further, because we have the Nielsen weekly top 100 going back to 2015, we construct measures of authors' past sales. We then use measures of the past sales of authors whose new books are in the top 100 this week. We implement this with the number of authors in the current top 100 whose previous sales are in one of seven intervals.

This gives us 17 possible instruments. To avoid choosing among them arbitrarily, we use the variable selection approach of Belloni et al. (2014). We estimate IV regressions in which we use LASSO techniques for the choices of a) which week dummies to include in the main equation, and b) which instruments to include in the first stage. The procedure selects 4 of the 17 possible instruments and 16 of the possible week dummies. Not surprisingly, the weeks before Christmas are selected. The resulting estimate of σ is 0.373 (se=0.0557).
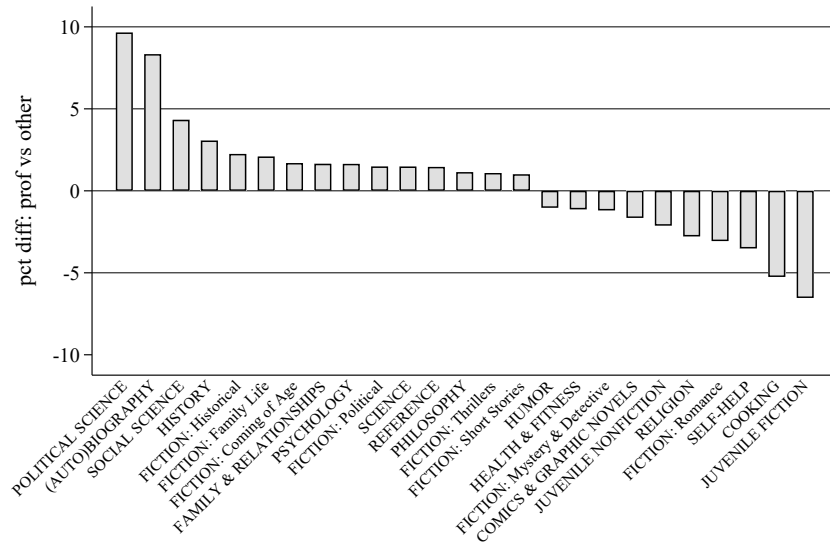
D   Additional Figures

Figure A.3: Amazon sales rank evolution of a sample book
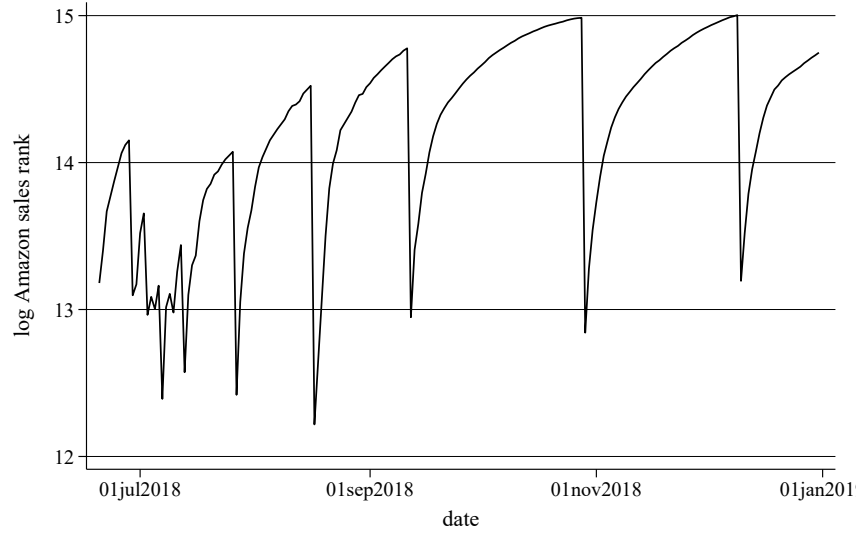
Notes: Daily Amazon log sales rank data for A Reaper at the Gates, by Sabaa Tahir. Large improvements show a day with a sale, followed by days of upward drift until the next sale.

Figure A.4: Composition of genres – reviewed vs. not

Notes: This figure reports the difference in the genre distributions between the professional outlets and others: the share of each genre in the professionally reviewed sample minus the share of the same genre in the remaining books. We include only the genres that differ by at least one percentage point.