Word of mouth

Stephan Seiler
Stanford Graduate School of Business

Song Yao
Carlson School of Management, University of Minnesota

Georgios Zervas
Boston University Questrom School of Business

November 21, 2017

1 The Impact of Reviews on Consumer Demand and Firm Pricing

Academic interest in consumer reviews goes nearly as far back as the appearance of the first online reputation systems on e-commerce platforms like eBay and Amazon. Reviews – the information units of reputation systems – typically consist of a numeric score (typically represented as a rating between 1 and 5 stars), the review submission date, and some open-ended text that allows consumers to evaluate a business (or a product) in their own words. Some reputation systems allow consumers to attach pictures to their reviews, and to separately rate businesses on dimensions such as service and price. Another commonly implemented feature lets consumers rate others' reviews as "helpful" or "useful". Reputation systems also display certain basic facts about reviewers such as their review history, a profile picture, and location, though they do not require that reviewers disclose their true identity. Most reputation systems allow anyone to submit to review, which has raised concerns about review fraud (**mayzlin2014promotional**; **luca2016fake**). Various strategies are in place to mitigate this concern including publishing reviews only by consumers whose purchases can be verified, highlighting reviews that can be linked to verified purchases using special badges, and relying on fraud-detection algorithms to discard fake reviews.

1.1 Review valence

The most common way in which reputation systems aggregate the plethora of information they collect is to compute an average rating for each business.[1] Average ratings, which are meant to capture overall quality, are prominently displayed and used to rank products and businesses in response to user queries. For instance, the query "hotels in San Francisco" on TripAdvisor is likely to return higher-rated hotels as the top search results.

Maybe due to wide use and simplicity, average ratings have been well-studied. By now, it is well-established that average ratings have a substantial causal impact on demand and pricing, though the magnitude of these effect varies by the timing and context of the study.

eBay's reputation system, which allows buyers and sellers on the platform to rate each other, was among the first to be studied and has been the focus of tens of studies (**ba2002evidence**; **houser2006reputation**; **lucking2007pennies**; **eaton2002value**; **bajari2003winner**; **kalyanam2001r** **mcdonald2002reputation**; **cabral2010dynamics**; **dewally2006reputation**; **jin2006price**). The findings of these papers, which rely on different methods and study different eBay product categories are broadly consistent: highly-rated sellers attract more bidders in their auctions, fetch higher prices, and sell their items with higher probability. Interestingly, **jin2006price** show that the while high-rated sellers can charge a premium, this is not because they sell higher quality products.

Similar effects for review valence have been demonstrated on other review and e-commerce

[1] An average rating is simply the *unweighted* mean of all individual ratings a business or product has received, often rounded to the nearest decimal point of half-star.

platforms. For example, **chevalier2006effect** find that the sales ranks of books on Amazon depend on their average ratings. Looking at Yelp, **luca2016reviews** and **anderson2012learning** estimate the impact of average ratings on restaurant demand and respectively find that a one-star increase in ratings causes a 5% increase in revenue and a 50% increase in the probability of being sold out. To demonstrate causality, both of these studies rely on a clever regression discontinuity (RD) design that exploits the fact that Yelp rounds average ratings to the nearest half-star: a business with a 4.24 average is rounded down to 4 stars, while a business with a nearly identical rating of 4.25 stars is up to 4.5 stars. Using the same RD strategy **luca2013digitizing** study ZocDoc, a platform that allows consumers to rate doctors and make appointments with them, and find that a half-star increase in ratings is associated with 10% increase that an appointment will be filled.

While many studies have investigated the impact of average ratings, less is known about other aggregate measures. One exception is the work of **sun2012variance** who studies the variance of ratings. Reputation systems do not explicitly display rating variances. However, most display histograms of reviews counts broken down by rating. By looking at these histograms consumers can infer rating variance. Motivated by a theoretical model, **sun2012variance** presents empirical evidence that increased variance in ratings has a positive impact on the demand of low-rated products. The key intuition behind this result is that low-rated products with some high individual ratings (*i.e.*, high variance) may be appealing to certain consumers, while products with consistently low ratings (*i.e.*, low variance) are less likely to be appealing to anyone.

A number of papers have examined the moderators of the relationship between ratings and demand. A common theme that has emerged is that the relationship between ratings and demand (or prices) depends on consumers' prior quality uncertainty. For instance, **luca2016reviews** finds that that Yelp ratings do not matter for chains because there exists little a priori uncertainty about their quality. Similarly, **lewis2016welfare** find that the impact of TripAdvisor ratings is much smaller for chain-affiliated hotels than it is for independently operated properties.

The effects we have discussed above are unlikely to represent long-term equilibrium outcomes. As consumers pay more attention to the increasing number of reviews accumulating online, the influence of these reviews on demand and pricing will also likely become stronger. **lewis2016welfare** provide some evidence for the evolution of these treatment effect by looking at hotel demand as a function of TripAdvisor ratings over a decade. The study finds that the impact of a 1-star increase in a hotel's average rating went for zero to 25% between 2004 and 2014.

In conclusion, we point out that the existence of a large number of studies of ratings across different contexts has enabled meta-analyses of their effects (**floyd2014online**; **babic2016effect**). By comparing treatments effects estimated in different settings, meta-analysis can reveal interesting moderators of the effect of ratings. For instance, **babic2016effect** find that early reviews are more important for new products than they are for new services.

1.2 Review volume

Another salient and well-studied feature of reputation systems is review volume. Review counts are typically displayed prominently next to average ratings, and consumers can also use them to infer quality. All else equal, we may expect that consumers will have less uncertainty about the quality of products with more reviews. Testing the hypothesis that changes in review volume affect demand raises a unique identification challenge: while more reviews may cause sales, the reverse is also true, *i.e.*, sales can result in more reviews. Thus, to avoid this reserve causality issue, reliable estimates of review volume on sales have primarily relied on experimental methods.

Two experimental studies investigate the link between review volume and demand. **resnick2006value** compare new eBay sellers with sellers that have an established reputation and find consumers' willing to pay is 8% higher for the latter. **pallais2014inefficient** studies oDesk, an online marketplace where employers can hire workers, and finds that workers who (randomly) received detailed feedback had better employment outcomes, including higher wages and earnings. An interesting implication of these studies is that having at establishing a reputation by having at least one review matters more than receiving an additional review.

1.3 Review text

Review text is inherently high-dimensional and typically not amenable to analysis by the same methods researchers use to study variables like review valence and volume. Therefore, most analyses of review text rely on a pre-processing step that transforms text into a small number of variables that capture variation along dimensions of interest. These variables can subsequently be analyzed using traditional econometric approaches.

At a high level, two pre-processing approaches have been employed in the literature to convert text into a manageable number of variables. First, researchers have used statistical approaches that rely on dimensionality reduction algorithms to map text onto low dimensional measures. Statistical approaches themselves vary in their sophistication from the direct application of simple formulas to compute metrics like readability to complex methods like topic modeling that can extract latent structure and meaning from unstructured data. **moe2017social** provide a comprehensive review of the statistical approaches used to analyze text from reputation systems and social media. The second approach is grounded in consumer behavior theory and starts by forming a hypothesis regarding the effects of specific text constructs. The presence of constructs in text is often detected manually (*e.g.*, using human coders), or via simple pattern matching rules.

buschken2016sentence. Write a paragraph about text. Cite **ghose2012designing**, **packard2017language**.

ghose2011estimating should be cited too.

ghose2012designing use text-mining study various aspects of review text on TripAdvisor, finding that simpler and shorter reviews that do not contain spelling errors have a positive impact on hotel demand.

packard2017language and **kupor2017spontaneous** offer examples of the second approach. **packard2017language** combine lab experiments and field data to investigate how the language consumers use in to endorse products affects how persuasive their reviews are. The authors find explicit endorsements ("I recommended this book") are more persuasive than implicit endorsements ("I enjoyed this book"). Interestingly, the authors find that in practice novice users are more likely to use explicit endorsements that experts, suggesting that novice reviews are more persuasive.

kupor2017spontaneous also use lab experiments to investigate the effects of endorsement authenticity. The study find that endorsements that are perceived to be more authentic are more convincing than endorsements that are perceived to be more thoughtful. For example, product endorsements that contain typographical errors are perceived as more authentic, thus enhancing their persuasiveness.

Both approaches for studying text are valuable. Statistical approaches can uncover patterns in text whose importance might not be evident a priori. At the same time, statistical approaches offer no direct connection to theory. Linking patterns uncovered by statistical methods to theoretical constructs can be difficult. Thus, consumer behavior theory can guide our search in the high-dimensional space of text patterns.

2 Other Literature Reviews

GZ: This para may be better suited for the overall chapter conclusion than this particular section. Discuss **dellarocas2003digitization** and **luca2015user**.