Review credibility as a safeguard against fakery: the case of Amazon

Wael Jabr

Supply Chain and Information Systems, Pennsylvania State University, University Park, USA

ABSTRACT
Online reviews remain a reliable source for customers when making purchase decisions. Yet, the pervasiveness of fake reviews jeopardises this reliability and questions the quality of this content. In this paper, we provide empirical evidence from a major online retailer that mitigation against fakery can be successful. To that end, we proposed, tested, and validated an approach, based on existing safeguards, to quantify the credibility of reviews and thus reliably reduce product uncertainty. We also showed that reviews with sufficient credibility signals were effective at influencing product sales, and this influence was prevalent for both niche and new products on the market. As such, this study offers a novel approach to mitigate the impact of fakery in reviews posted to online infomediaries. Our work focuses primarily on Amazon as a major retailer but also provides further support by drawing on Yelp, another major review platform.

1. Introduction

The otherwise thriving online sales channels have their own unique – albeit sometimes limiting – characteristics, including the inability of customers to touch and handle the product. This setup engenders some uncertainty about product quality and features. It also makes user reviews an essential component of the success of these online channels now also acting as infomediaries. The premise is that reviews provide prospective customers with information – supplied by previous customers – to help them in their purchase decisions.

This setup worked flawlessly for a long time, as evidenced by the rich academic and professional literature on the impact of user reviews on sales and satisfaction. The unintended consequence seems to be, however, that these reviews became subject to intentional manipulation – also referred to as fraudulent, fake, spam, or chill reviews; we hereafter use the term “fakery.” This phenomenon has since become rampant – or so the extensive coverage in the popular press would have us believe; see Appendix I for anecdotal evidence. The worry is that manipulation of reviews could become severe enough to bias the information they provide and consequently undermine trust in the infomediary.

In response, and to safeguard the review system, research has attempted to quantify the bias resulting from fakery. In this vein, the last few years have witnessed an increasing interest in detecting review manipulation. This task, however, is not a trivial one. A multitude of approaches have been proposed in the computer science and information science literature to detect such reviews (Kumar et al., 2018; Shukla et al., 2019). Unfortunately, these approaches suffer from a fundamental limitation, that of the lack of a rigorous validation of their classification performance. This shortcoming is attributable to either: the lack of a gold standard readily available to researchers, possessing True Positives and True Negatives – that is, reviews known to be fake and reviews known to be authentic, respectively – as in Mukherjee et al. (2012); or a lack of replicability, due to approaches that are tailored solely to one infomediary and to a specific set of reviews, as in Kumar et al. (2018). Alternatively, rather than detecting and quantifying fakery in online reviews and subsequently correcting for bias, a recent stream of research attempts to study its impact, or lack thereof, on sales (Luca & Zervas, 2016), business visibility (Lappas et al., 2016), evaluation of physicians (Lu & Rui, 2018) and customer trust (Ananthakrishnan et al., 2019).

In this paper, we proposed an alternative information-based approach to quantify the credibility of product reviews. This approach leverages features implemented by the respective hosting platforms that are intended to help in the vetting of the reviews, such as verified purchases and helpful votes in the case of Amazon. Here, anchored in pertinent literature, credibility is defined as the extent to which the review information provides signals perceived to be believable and factual. We also provided a validation of this approach. We then proposed an econometric approach to investigate the potential economic impact of these credibility signals.
We contribute to the literature on information bias (e.g., Mayzlin et al. (2014)) by reframing the problem of fakery as a problem of bias and framing the objective of this research as aimed at safeguarding the review system. Specifically, this paper provides two fundamental research venues: a method for formally quantifying the credibility of information and an approach to evaluate the influence of the corresponding signals on the functioning of online markets.

To answer the first research question, we proposed an approach for assessing the credibility of online content. By doing so, we framed the fakery problem as one of quantifying credibility amidst prevalent fakery and analysed its impact. To that end, we developed an approach for quantifying the credibility of product reviews, relying on featuring tools that the infomediary provides. Through the lens of information theory, we showed how, over time, these tools enabled us to reach an information credibility steady state from which customers were able to derive credible information about products they were interested in. We also provided validation of this approach at two major infomediaries.

To answer the second research question, we then studied the economic impact of posted reviews and showed that the credibility signals provided by reviews are influential in swaying users’ opinions. The paper therefore shows that, within the framework analysed, reviews mediated by the infomediary have enough checks and balances to be shielded against potentially damaging manipulation bias.

2. Literature review

Three literature streams are relevant to this work: the influence of reviews on product performance; the impact of bias in online settings, including the role of fake reviews; and the credibility of information in online settings.

2.1. Influence of reviews on product performance

The growing body of literature on online reviews spans a wide spectrum of disciplines, including Information Systems, marketing, and economics; several literature syntheses are now available, including Floyd et al. (2014), King et al. (2014), and Babić Rosario et al. (2016). This literature has highlighted the influence that user reviews have on sales, quantifying increases in demand (Li et al., 2011). This demand effect has predominantly been evaluated using two metrics: review volume – i.e., the total number of reviews a product receives – and review valence – i.e., the rating a user gives the product (Archak et al., 2011; Forman et al., 2008; Jabr & Zheng, 2014; Watson et al., 2018).

Subsequent studies have investigated related metrics such as helpful votes (Dhanasobhon et al., 2007), rating variance (Sun, 2012), review informativeness (Archak et al., 2011), and product and consumer characteristics (Zhu & Zhang, 2010). Additionally, research has shown that different types of consumers benefit from the variety of informational value that reviews offer (Chunhua et al., 2015; Sun, 2012). We built on this body of research which studies the association between product performance, essentially in the form of sales, and the posting of reviews, essentially through their volume or valence. The model we develop rather focused on credibility and its economic impact while at the same time accounting for their intricate econometric relationship.

2.2. Impact of biases in online settings

While the role of reviews has been well-researched and established, these reviews have been shown to include a certain degree of bias that may – intentionally or not – influence customers and ultimately deceive them. In fact, exploring the dynamics of user reviews, researchers found an increasingly negative rating trend over the lifetime of the product (Godes & Silva, 2012; Li & Hitt, 2008; Moe & Trusov, 2011). Subsequent work pinned this trend on a herding behaviour of reviewers (Lee et al., 2015) and on the extremity of reviews (Nagle & Riedl, 2017).

At the same time, the literature has identified the predominance of positive reviews. Zervas et al. (2015) summarised several underlying drivers: prior ratings inadvertently biasing the posted opinions of subsequent reviewers (Godes & Silva, 2012; Godinho De Matos et al., 2016; Qiu et al., 2017); under-reporting of negative reviews, where reviewers fear user base retaliation (Dellarocas & Wood, 2008; Fradkin et al., 2018) or just do not feel the need to review (Chen et al., 2016); and self-selection, where customers who write reviews are more likely to be early adopters who are satisfied with the product (Li & Hitt, 2008). Eventually, these biases came to be classified as confirmation bias, self-selection bias, directional bias, under-reporting bias, acquisition bias, herding bias, or deception bias; see Hu et al. (2017) for an overview.

A nascent literature aims to quantify this bias in consumer opinions and propose approaches to correct for it. Chen et al. (2016), for example, provided a method to model the data-generating process of reviews and then rectify the underreporting bias through an inverse probability weighting.

In that same vein of research, fakery is yet another type of bias that skews ratings. Kumar et al. (2018) and Ananthakrishnan et al. (2019) quantified it in the context of Yelp, and Shukla et al. (2019) quantified it in the context of health forums. Although fakery is tricky to define (Peltzman, 1981), it is typically characterised by...
misrepresentation of the attributes of the promoted product (Nagler, 1993) and, consequently, misrepresentation of the corresponding product’s expected utility. In fact, both the Commission of the European Communities’ directive on misleading advertising (European Union, 1984) and the U.S. Federal Trade Commission Act (Federal Trade Commission, 2008) regulate this space, prohibiting “unfair or deceptive acts or practices,” including both misstatement of facts and failure to disclose important information that consumers should know.

However, very little is known of the extent of deceit online. To our knowledge, the only formal, reliable, and large-scale attempt to quantify deceit was conducted by the UK’s Competition and Markets Authority in 2015 and showed that its prevalence is quite limited (see excerpts in Appendix II), with only 4 to 8% of respondents qualifying their experience as “a little worse” or “much worse” than the impression the online reviews had given them (Competition and Markets Authority, 2015).

Furthermore, research has shown that consumers respond only weakly to exaggerated claims, and hence deception ends up yielding very little benefit for the promoters (Lee et al., 2014; Saulnier, 2015; Zinman & Zitzewitz, 2016). Infomediation control mechanisms have also been shown to help mitigate the impact of deceit (Luca & Zervas, 2016; Mayzlin et al., 2014). We built on this emerging body of literature to provide a novel approach for the assessment of fakery’s impact on the online market.

2.3. Credibility of information in online settings

Information credibility is a complex concept and is the subject of a long-standing body of research. In their influential 1951 work, Hovland and Weiss adopted the lens of the receiver of the information to emphasise the perception element of credibility. As such, credibility is typically considered a subjective perception (Fogg et al., 2001; Gunther, 1992) determined by the receiver’s judgements (Kaye & Johnson, 2011). Nevertheless, various elements have been identified as contributing to credibility, including perceived truthworthiness; perceived expertise of the information source (Hovland & Weiss, 1951); believability, trust, and perceived reliability of that source (Self, 1996); believability and factuality (Cheung et al., 2009); trust in reviewers (Chen & Lurie, 2013; Schlosser et al., 2006); and trust in reviews’ text and ratings (Korfiatis et al., 2012; Yin et al., 2016). Lazer et al. (2018) proposed that infomediators incorporate such elements in order to provide customers with indicators of content quality in online settings.

Researchers agree that credibility formulation is realised as a multi-dimensional concept that is captured by multi-factor measures. Prior studies have attempted to develop models for characterising information credibility. One such well-structured model is the Yale attitude change model (Hovland et al., 1953), which suggested that four factors affect information persuasiveness. These consist of the source of the content, through the corresponding level of expertise and trustworthiness; the message, through content quality, sidedness, and consistency; the medium, in our case the review; and the receiver, through any prior knowledge. More recently, in the same vein, but in the context of online content, Edelman (2017) indicated that, “questions of review authenticity and trustworthiness recede into the background when a review provides all the information needed to confirm its assessment” (p. 642). He identified the following criteria as adding to the review credibility: information confirming its authenticity and trustworthiness, such as photographs or reference to some source beyond dispute; a vibrant community; and verifiability of purchase. Luca (2017) also wrote that, to enhance the credibility of reviews, platforms can give greater weight to trustworthy reviewers with longer transaction histories. An extensive literature in social psychology and related areas has shown the importance of credibility in influencing the impact of a persuasive message (Cialdini, 1985; Chen et al., 2008).

Indeed, a variety of review platforms have been actively trying to improve the credibility of their content by means of a variety of validation artefacts. For instance, Expedia implemented a review verification mechanism that was shown to increase the cost of posting a fake review and thus acted as a deterrent, making promotional reviews less common on Expedia than on TripAdvisor, where no verification mechanism was imposed (Chevalier et al., 2018). Similarly, Amazon only labels reviews as “Verified” when the reviewer purchased the product being reviewed through the platform and after the reviewer has spent more than 50 USD on purchases. Amazon also allows review readers to indicate that review was helpful to them by casting a “Helpful vote.” In this research, we expanded on this literature on credibility to develop the perception of information credibility, derived from information that Amazon provides next to posted online reviews, and then validated in the context of Yelp.

3. Data

To gather the data for this research, we first selected a total of 1,000 products listed on Amazon.com, belonging to four product categories: Automotive; Home & Kitchen; Health & Personal Care; and Grocery & Gourmet Food. These categories were chosen to ensure some variability. The selected products
were classified as follows. For each of the four aforementioned categories, we selected ten sub-categories – for instance, one such sub-category under the category Grocery & Gourmet Food was Cooking & Baking. From each subcategory, we selected an average of 85 products to include in the final product sample set. The mix of individual products to be tracked was characterised by variability in the dimensions proposed in the literature: product involvement, product benefit, and frequency of purchase, as documented in the meta-analysis by Floyd et al. (2014) and the work by Gu et al. (2012); and search/evaluation efforts and risk reduction, as documented in the review by King et al. (2014). As such, the assembled sample had a reasonable level of variability in the number of reviews received, sales rank, and price, as well as utility and hedonic qualities. We tracked those products recurrently using scripts developed in Java over a period of one year. Table 1 presents an overview of the data collected, including the quantity of products and their respective number of reviews for each of the category types. Since we collected the same reviews recurrently over time in order to extract the time-variant helpful votes, the total number of reviews with duplicates included is presented in the last column. Appendix III provides dataset statistics.

For each product in our set of tracked products, we collected information about the product, its reviews, reviewers, and number of helpful votes every two weeks between mid-February and mid-November 2015. Specifically, we collected static, as well as dynamic, data related to these products. For the static components, data collection included product description, including ID and release date, and for each review, we collected the review post date, the review text, and the reviewer name and ID. For the dynamic components – that is, that changed over time – data collection included price, sales rank, average rating, individual reviews received, and their corresponding helpful votes. We also collected the featured reviews that the retailer selects to present to customers and displays prominently on the front page. The complete dataset therefore consisted, by the end of the collection period, of a little over one million unique reviews posted for over one thousand products and the corresponding dynamics. Given the recurrence of the data collection, the complete dataset consisted of more than 17 million reviews collected repeatedly over the different collection dates; see Table 1. The crawlers collected information posted on the product page, including the subset of helpful reviews listed under the heading “Most Helpful Customer Reviews,” as illustrated in Figure 1.

4. Methodology

We started by presenting and validating the information-based approach for quantifying the credibility of reviews, then proceeded to quantify the economic influence of the resulting credibility signals to mitigate fakery.

4.1. Information-based approach to credibility

We first developed an approach for ascertaining the information provided in reviews and quantifying its credibility. The rationale behind this approach is that, as information from reviews with various levels of credibility starts accumulating over time, prospective customers may increasingly trust the credibility of the information and shop with greater confidence. This approach was grounded in the literature on the quantification of credibility, including the Yale model (Hovland et al., 1953) and its newer adaptations in the online review space.

More specifically, the nature of the online review infomediary is one where the true quality of the product is not readily available to customers. Additionally, the information available might be limited, conflicting, or fake, with few ways to gauge its credibility. As such, there is a need to better determine the quality of review information generated about products. Within these limitations, customers attempt to infer the credibility of products’ review information.

While the true quality of a product may not be directly observable at first glance, review information is observable. Over time, these reviews gain helpful votes that better signal their credibility. The accumulation of votes generated through an exogenous process results in Amazon featuring those reviews. We exploited this setup to derive our model, as detailed below.

4.1.1. Model

Following well-established literature (see, for example, Mullainathan and Shleifer (2005), Xiang and Sarvary (2007) and Zhu and Dukes (2015)), we modelled the process of arrival of featured reviews as a data string $S$. This string consisted of a series of featured helpful reviews $(r_{i}^{\text{featuredhelpful}})$ that individually and collectively provide a credible string of

| Table 1. Distribution of Products Across Various Categories. |
|-------------|-------------|-------------|
| Category Type       | Product Count | Review Count | Review Count Across Study |
| Automotive          | 176          | 178,665     | 2,986,577                 |
| Grocery & Gourmet   | 290          | 161,562     | 2,639,333                 |
| Food                | 272          | 420,323     | 6,476,675                 |
| Health & Personal   | 262          | 329,484     | 5,435,760                 |
| Care                |              |             |                          |
| Home & Kitchen      | 1,000        | 1,090,034   | 17,538,345                |
| Total               |              |             |                          |
information about the product investigated, thus $\text{CredibleString} S = \left( r_i^{\text{featured helpful}} \right)_{i=1}^R$.

These featured helpful reviews are independent and identically distributed (i.i.d.) draws from a gamma distribution with parameters $(s_i, \lambda)$ where $s_i$ is the shape parameter and $\lambda$ is its scale parameter. It is worth mentioning here that the i.i.d. assumption was not required for the overall arrival of reviews, which one can easily assume is not independent, as shown by the extant literature, due to herding and self-selection among other factors (Li & Hitt, 2008). Rather, this assumption just pertained to the featured helpful reviews that formed the string $S$. Those featured helpful reviews were eventually randomly assorted into the string $S$ irrespective of their temporal posting sequence or their authors, but rather based on the helpful votes over time and featuring by the platform. Thus, any given review of the string $S$ in position $i$ was credible with probability $p_i$. Furthermore, $p_i$ got updated frequently as additional credibility signals for individual reviews – in the form of review type and helpful votes – accumulated over time, thus adding a time suffix and turning it into $p_{ii}$. Initially, at the time of its posting, each review fell into one of three categories: a review that was verified by Amazon to have been posted by a purchaser; a Vine review posted by a customer that received the product directly from Amazon on a trial basis; or neither Verified nor Vine.

Over time, and irrespective of their type, reviews received helpful votes that increased their credibility signals and prompted further increases of $p_{ii}$. The probability of credible information revealed by a review about a product $P(\text{CredibleString} S)$ became the probability of the aggregation of each of the credible reviews, $r_i^{\text{featured helpful}}$, that a product received: $P(\text{CredibleString} S) = P \left( \sum_{i=1}^R r_i^{\text{featured helpful}} \right)$.

With the assumption that featured helpful reviews arrived independently of each other – that is, they were posted on the webpage independently – and that helpful votes arrived independently as well, the probability of a credible review was evaluated as a gamma-distributed random variable with parameters $(s_i, \lambda)$. The String $S$ of reviews $r_i$ therefore became a gamma-distributed variable itself with parameters $\left( \sum_{i=1}^R s_i, \lambda \right)$ as the sum of gamma-distributed variables. A sketch of the analytical proof is presented in Appendix IV. Gamma distribution was an adequate choice to model heterogeneity in the accumulation of helpful votes, as well as heterogeneity in the type of reviews – Verified, Vine, or neither. In fact, and closely related to the setting at hand, gamma distribution has been widely used to model heterogeneity in purchasing propensity among customers (Moe & Fader, 2004; Zhiqiang et al., 2012). In distinct contexts,
gamma distribution has also been used to model the size of loan defaults, the aggregate severity of insurance claims, and the accumulation of water flow (Mun, 2008).

Furthermore, gamma distribution is a flexible, skewed distribution with a density that increased from the origin to a mode at \((s_i - 1)/\lambda\), then decreased for \(s_i > 1\). For each random variable \(r_i\), accumulating helpful votes, varying \(s_i\) allowed for varying the skewness and thus, in our context, the count of helpful votes received.

Next, we derived the parameters of the gamma distribution \((s_i, \lambda)\). For that, we started with the mean and variance of the distribution. It can be easily shown that the mean \(E[r_i]\) and variance \(Var[r_i]\) of the distribution are be expressed as \(E[r_i] = s_i \times \lambda\) and \(Var[r_i] = E[r_i^2] - (E[r_i])^2 = s_i \times \lambda^2\). Thus, \(s_i = (E[r_i])^2 / Var[r_i]\) and \(\lambda = Var[r_i] / E[r_i]\), with common \(s_i\) and \(\lambda\) across all product reviews.

### 4.1.2. Parameter estimation

To improve the quality of the parameter estimation, while at the same time avoiding overfitting, we added a generalisation step that expanded the identification process of the parameters beyond an individual product and accounted for all the products with similar characteristics. Thus, rather than parametrising at the product level, we conducted this process at the category level. The intuition belief was that reviews posted to similar products were drawn from a similar distribution of helpful votes accumulation, at least vis-à-vis the credibility dimension.

We therefore started by identifying products with similar characteristics. To that end, using the product title and textual description listed on Amazon.com, we ran product clustering using SAS Enterprise Miner to group products into clusters of closely related characteristics. To ensure the validity of the clustering results, two graduate research assistants separately reviewed and validated the classification outcomes. The results were then compared for consistency between the two research assistants. Given the manageable number of products, we ensured that the process was repeated until we achieved full agreement.

So, in sum, for each product belonging to a given cluster, the corresponding parameters of the gamma distribution were derived as detailed above across all their reviews. Then a grand average of all of \(s_i\) and \(\lambda\) in a given cluster was derived. We ran tests of fitness of the gamma distribution for each cluster’s reviews using a Wald \(\chi^2\) test.

### 4.1.3. Model estimation

With the parameters of the gamma distribution derived, we estimated the credibility of product information based on the aggregated credibility of its corresponding reviews. We then tracked the evolution over time of this credibility. Figure 2 depicts – for four representative products, identified as Sample Products 1 through 4 – the change in the product’s credibility over time as additional helpful votes were added to reviews. The selected products were representative of the various ways that information credibility improved over time. It is worth noting that the individual products depicted had different starting points for credibility, which depended on the volume of accumulated helpful votes at the start of the data collection. Sample Product 1 showed a dramatic change in the credibility of the information about that product over time based on accumulated helpful votes. In the first couple of weeks, its credibility increased from close to 0 to close to 0.5 and then rose steadily after that until the end of the two-year data collection period. On the other hand, sample Product 4 displayed rather slow improvement in the credibility of its information, which can probably be attributed to the fact that, by the time data collection began, the product’s information had    

![Figure 2. Change in Probability Distribution of Product Credibility over Time (Amazon).](image-url)
already garnered an adequate credibility level of over 0.5.

To summarise, the analysis focused on proposing an alternative measure, namely, credibility of information, to assess the extent to which customers’ need for credible product information was met. Here also, and unlike the extant literature, the proposed measure circumvented fakery-related uncertainty regarding the quality of information.

4.2. Validation

We next turned to further validating the applicability of our proposed approach in a different setup, where non-credible reviews were explicitly labelled by the infomediary platform. In fact, the review site Yelp labels some reviews as “not recommended” when the site deems them not to be credible. According to Yelp’s stated policies, this labelling is based on: the credibility of the reviewers, such as reviewers about whom very little information is available; the content of the review, as in unhelpful raves or rants; or the source, for example, reviews originating from the same computer or authored by friends of the business (Yelp Support Center, 2016).

Along with this binary labelling of “not recommended,” Yelp allows reviewers to indicate whether they checked in at the business – comparable to a validated purchase at Amazon – and allows readers to post helpful votes on reviews, like helpful votes at Amazon.

The dataset consisted of 7,409 businesses, 439,621 reviews, and an additional 67,760 “not recommended” reviews that constituted 13.35% of the reviews. Following a derivation approach similar to that above to identify the parameters of the credibility distribution, Figure 3 shows how non-credible reviews suppressed credibility signals, thus requiring a longer time to reach a plateau and attain a high level of credibility for a given business. However, after that, there was no discernable impact from non-credible reviews.

4.3. Economic analysis

The second step in this analysis was to quantify the impact of credibility signals on product performance. The objective was to identify whether products that provided credibility signals by means of credible reviews ended up benefiting in terms of those products’ market performance. We also made use of a variety of cues displayed on individual product pages on Amazon.com intended to further help customers reduce their uncertainty about a product they were considering. These consisted of the total number of reviews received and their average ratings – as a proxy for a vibrant community – and a featured set of the most helpful customer reviews, sorted by helpfulness, as a proxy for information confirming authenticity and trustworthiness. With the data collection tracking product pages recurrently, we were able to identify the reviews that were displayed on product pages under the label “Most Helpful Customer Reviews” each time the data was collected.

4.3.1. Estimation

As is customary in the Information Systems literature, the analysis investigated the impact of product characteristics and posted reviews on product sales performance, with sales rank as a proxy. Accounting for the dynamics of this setup enabled us to overcome biases in the estimation of main effects. Arellano and Bover

Figure 3. Change in Probability Distribution of Product Credibility over Time (Yelp).
(1995) followed by Blundell and Bond (1998) devised an approach to overcome these identification challenges through the use of a dynamic panel model with robust standard errors. As detailed in the literature – for example, Roodman (2006) and Bapna et al. (2013) – this approach provided unbiased and efficient estimators based on moment equations constructed from further lagged levels of the dependent variable and the first-differenced errors.

This approach was also designed for situations such as the one at hand. First, the approach is characterised by linear functional relationship, as is the case between a product’s reviews and its rank in our setting. Second, the approach is characterised by a dependent variable that is both dynamic and correlated with its past realisations. This was appropriate for the current setup, as the behaviour of the dependent variable Sales Rank was best explained with both contemporaneous and past values of its own, as well as other explanatory variables. Third, the approach is characterised by several independent variables that are not strictly exogenous, that is, correlated with past and possibly current realisations of the error term. This was the case with the current setup for variables like Total Reviews Received, Average Ratings and several others that were obviously not strictly exogenous. Fourth, the approach deals with situations with fixed individual effects. In the current setup, those individual effects consisted of products’ inherent characteristics that did not change over time, such as shape and colour. Fifth, the approach is characterised by panels with heteroskedasticity and autocorrelation that can be readily tested for. Appendix V provides further theoretical background.

We now proceed to describe the variables for the dynamic panel model. The independent variables of interest consisted of the vector of variables related to helpful reviews posted on the product page (\(\text{RW}^h\)). This vector included the average rating, standard deviation of ratings, count of featured reviews, and credibility. The control variables consisted of review-related variables (Total Review Volume and Overall Rating) and a product-related variable (Price). We used a two-step GMM estimation, as the estimates were more asymptotically efficient than one-step when using the Windmeijer (2005) correction procedure and robust standard errors. Table 2 provides a listing, description, and summary statistics of the various variables.

In its most general form, the regression was as follows:

\[
\text{SalesRank}_{it} = \text{Credibility}_{i,t-1}\alpha + \sum_k \text{RW}^h_{i,t-1}y_t + \text{RR}_{i,t-1}\delta + P_{i,t-1}\beta + \text{SalesRank}_{i,t-1}\rho + \varepsilon_{it}^k
\]

(Equation 1)

where \(i\) – index for the product cross section & \(t\) – the index for data collection timestamp, \(h\) – index for helpful reviews \(\text{Credibility}_{i,t-1}\) – probability measure of product credibility, \(\text{RW}^h_{i,t-1}\) – vector of helpful review-specific fixed effects regressors consisting of [Avg. Helpful Rating, Std. Dev. Helpful Rating, Total Featured Helpful] \(\text{RR}_{i,t-1}\) – vector of overall review-specific fixed effect regressors (Overall Rating & Total Reviews) \(P_{i,t-1}\) – price, a product-specific fixed effect, \(\alpha, \rho, \beta, y\) and \(\delta\) – regression coefficients, \(\varepsilon_{it}^k\) – idiosyncratic error.

The estimation results are presented in Table 3. We start by estimating Equation 1 using the full dataset in Column 1.1 of Table 3.

### 4.3.2. Additional specifications

With the objective of obtaining further insights, we then derived the estimation of Equation 1 on two subsets, one consisting of only niche products – resulting in Equation 1.2 – and another consisting of new release products – resulting in Equation 1.3 – to complement the analysis for the full dataset, Column 1.1. In fact, the academic literature has highlighted the issue of “niche” products – also referred to as long tail – as being a major challenge for infomediaries, as these products do not typically induce the generation of user content (Gu et al., 2012). We therefore

### Table 2. Variable Description and Summary Statistics.

| Variable                   | Description                                                                 | N  | Mean   | Std Dev |
|----------------------------|------------------------------------------------------------------------------|----|--------|---------|
| Log Sales Rank             | Amazon product sales rank (log transformed)                                 | 18,199 | 5.329  | 2.030   |
| Credibility                | Signals of how believable and factual reviews are; measured as a probability| 18,199 | 0.718  | 0.245   |
| Avg. Help Rating           | Average rating of reviews posted as helpful                                  | 18,199 | 4.145  | 0.843   |
| Std. Dev. Help Rating      | Standard deviation for rating of reviews posted as helpful                   | 18,199 | 0.889  | 0.606   |
| Total Featured Help Reviews| Total number of reviews featured as helpful on the product                   | 18,199 | 6.334  | 1.792   |
| Total Reviews              | Total number of reviews posted to the product                                | 18,199 | 1.670  | 2.259   |
| Overall Rating             | Average customer rating posted to the product                                | 18,199 | 4.373  | 0.316   |
| Log Price                  | Price posted (log transformed)                                              | 18,199 | 3.009  | 0.749   |
| All variables are measured for product i at time period t
Table 3. Regression Results for Equation 1.

| Equation 1 | (1.1) | (1.2) | (1.3) |
|------------|-------|-------|-------|
| Log Sales Rank | All products | Niche | New |
| L.Credibility | $-0.318^{**}$ | $-0.313^{***}$ | $-0.316^{***}$ |
| L.Avg Help Rating | $(0.184)$ | $(0.114)$ | $(0.0866)$ |
| L.StdDev Help Rating | $-1.647^{*}$ | $-1.567^{*}$ | $-0.509$ |
| L.Total Featured Help | $(1.271)$ | $(1.078)$ | $(1.370)$ |
| L.Total Reviews | $-0.015^{**}$ | $-0.020^{**}$ | $0.221^{**}$ |
| L.Overall Rating | $(0.091)$ | $(0.263)$ | $(0.0741)$ |
| L.Log Price | $0.576^{***}$ | $0.511^{**}$ | $-0.727^{**}$ |
| L.Log Sales Rank | $-0.427^{***}$ | $-0.313^{***}$ | $-0.316^{***}$ |
| FE | Y | Y | Y |
| Category Dummy | N | N | N |
| Model fit statistics | | | |
| Wald $\chi^2$ | 1080 | 45.9 | 182.1 |
| Prob $>\chi^2$ | *** | *** | *** |
| Prob AutoRegressive | *** | *** | *** |
| AR(1) p < | NS | NS | NS |
| AR(2) p < | NS | NS | NS |
| Hansen test of overidentification | 100 | 100 | 100 |
| $\rho$ | NS | NS | NS |

L indicates lag of the variable; Standard errors listed in parentheses.
Significance listed at *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, & NS not significant.
Model estimated with fixed effects and includes category dummy.
Column (1.1) displays the results of Equation 1 with helpful-related variables run on the full dataset. Column (1.2) displays the results of Equation 1 with helpful-related variables run on the niche partition. Column (1.3) displays the results of Equation 1 with helpful-related variables run on the new partition.

We constructed a data subset to only include products that might be labelled niche. We labelled a product niche based on the volume of reviews received and its median sales rank. A product was identified as niche if it was in the lower 10% of the distribution of review volume – having a median rank > 7,000 over the data collection period – and in the long tail of the upper 90% of sales rank distribution – having a review volume < 300 by the end of the data collection period. We complemented this classification by manual check to ensure the validity of the resulting subset.

Another issue the academic literature has highlighted is the newness of the product and the resulting cold start. Research has shown the influential impact of user-generated content, be it offline or online, on such new products (Brooks, 1957; Dellarocas, 2006). We therefore constructed another data subset to only include products that might be labelled “new.” We labelled a product new if it had never received a review prior to the start of data collection and subsequently received at least 10 reviews. It is worth noting that we ensured that products considered in the new data subset were excluded from the niche data subset in case of overlap. Table 4 provides a summary of the various data subsets used.

We conducted two robustness checks vis-à-vis the validity of using Verified purchase and Vine reviews, as well as Helpful votes, as tools to conduct the analysis, as these measures provided additional signals of credibility. The findings are presented in Appendix VI.

### 4.4 Results

Table 3 shows the regression results for three different variations of our model using Helpfulness related variables: full dataset in Column 1.1, niche subset in Column 1.2, and new subset in Column 1.3.

We started by analysing Column (1.1). The coefficients for most of the variables in the regression were in line with the findings from the extant literature. For example, Overall Rating and Total Reviews were significant and negative, indicating that higher review ratings or higher review volume promoted lower rank and thus higher sales. Similarly, Price was significant and positive, indicating that higher prices correlated with higher rank and thus lower sales.

Other variables warranted further investigation. The coefficient on Avg. Helpful Rating in Column 1.1, which was significant and negative, indicated that the featured helpful reviews posted on the product page were influential in convincing customers to make purchases and thus improving the sales rank.

Together, these two findings indicated that the infomediary held a tight grip on review quality (through the various validation cues), and mitigated fakery issues through credibility and featuring approaches. This resulted in ensuring customers ultimately maximised their utilities and purchased products that were most suitable to their tastes and needs. This was in line with recent findings in the literature. In their study on deceptive advertising at ski resorts, Zinman and Zitzewitz (2016) found that consumers responded only weakly to exaggerated claims, and that, as a result, advertisers reaped few benefits from deception in equilibrium.

The other noteworthy finding was that the coefficient on Avg. Helpful Rating in Column 1.2, was not significant, indicating that the featured helpful reviews posted on the product page did not significantly move

| Table 4. Dataset and Subset Description. |
|------------------------------------------|
| Datasets | Description |
| Full dataset | Dataset with all 1,000 products tracked over two years |
| Subset 1 – Niche | Dataset with products gathering < 300 reviews and > 10,000 median sales rank by the end of the two-year data collection |
| Subset 2 – New | Dataset with products gathering no reviews at the initiation of the data collection & ultimately receiving at least 10 reviews by the end of the two-year data collection |
customers’ decision-making. In fact, the pool of featured helpful reviews may not have been deep enough to provide the needed information. This justification became even clearer, given that the coefficient for Total Featured Help was significant and negative, indicating that the longer the list of these reviews, the lower the rank. The coefficient on Credibility, however, remained significant and negative.

When investigating new products only, the coefficient on Avg. Helpful Rating in Column 1.3, was not significant, a further indication that customers were wise enough to realise that the subset of reviews posted on the product page had not “matured” long enough over time to be a good indication of the quality of the product. Total Reviews and Overall Rating were significant, indicating that these were the clues customers used to make purchase decisions. The coefficient on Credibility, however, remained significant and negative.

The role played by the infomediary was therefore quite central in funnelling credible information to customers despite – or maybe because of – the hype regarding the prevalence of fakery in review environments. In fact, this essential function is what infomediaries were designed for, namely, protecting buyers and sellers from the opportunistic behaviour of other participants by becoming an agent of trust (Bakos, 1998).

4.5. Discussion

At the infomediary we studied, current mechanisms – specifically the featuring of helpful reviews and helpful voting by subsequent users – seemed quite effective at signalling credibility and countering any potential dilution of information quality caused by fakery. As such, the infomediary was able to fully play its role in safeguarding the quality of the information presented. These findings were in line with analytical results by Glazer et al. (2019), who showed that, as long as a proportion of reviews was not fake, “the belief of every customer moves in the correct direction at every step.”

However, this setup did not seem as fitting for niche products, and the infomediary had to play a greater role in safeguarding the quality of information in those settings. This setup therefore required greater attention and diligence from the infomediary. This finding was in line with Zhu and Zhang (2010), who found that one negative review about a niche product could be detrimental because of the scarcity of available information and that in general online reviews were significantly influential. They added that, “niche producers should be more concerned about manipulations of online review systems because online reviews could significantly affect their sales.”

In truth, all markets require some level of credibility (Akerlof, 1970), and this is of utmost importance for online markets characterised by information asymmetry and product quality uncertainty. To that end, online markets have long played the role of infomediary, thus helping channel a multitude of information sources to the customer. Amongst these sources, online reviews have become an essential component of the marketing mix (Chen & Xie, 2008). To ensure their credibility, online markets have employed a variety of feedback-based systems (Dellarocas, 2003) that enable the building of customer trust and market efficiency. Such feedback-based systems have been studied extensively in a variety of markets, such as eBay, and have proven their usefulness. In this paper, we contribute to this body of literature by providing empirical evidence that Amazon’s role in curating and featuring reviews fulfils its intended purpose in contributing to trust in its online market.

This paper joins an emerging stream of research in providing empirical evidence regarding the crucial role of the infomediary in online settings. Like Mayzlin et al. (2014), who found that manipulation was reduced when the infomediary put quality differentiation mechanisms in place, this paper found that these differentiation mechanisms worked well to support customers. Furthermore, these mechanisms had additional benefits, including reducing information overload – as pointed out by Jacoby et al. (1974) in an offline setting, and more recently by Branco et al. (2016) in an online setting.

5. Implications

Our research echoes the recent call in the literature for infomediaries to implement credibility mechanisms. In fact, Lazer et al. (2018) called for a similar approach to the one we proposed, where the platform provides customers with signals of source quality that get incorporated into the delivery of content. This credibility approach was also found to be superior to approaches that censure fake reviews, with Glazer et al. (2019) cautioning against strict removal of fakery, as it leads to leaving behind fake reviews that are difficult to identify as such and that end up becoming more influential. As such, platforms that host reviews have a considerable stake in adding credibility signals to those reviews, rather than deleting suspected fake reviews. These signals could be either community-based, like helpful votes, or platform-based, such as Verified purchases. Additional signals might incorporate the trustworthiness of the reviewer – based on the number of reviews they had authored, for instance, as well as the informativeness of the content, for instance, using seals or badges of recognition from the platform.
6. Conclusion and future directions

Jeff Bezos was recently quoted as saying that, “we don’t make money when we sell things, we make money when we help customers make better decisions” (see his 2013 interview with Harvard Business Review’s IdeaCast). As such, Amazon and other infomediaries have every incentive to ensure their customers remain well informed.

One major threat that could potentially dilute the quality of this information the fakery manipulation in online reviews. In this paper, we demonstrated, in a two-step approach, that this problem is a mere fallacy. First, we showed that even if fakery existed – and it might – customers and infomediaries are well equipped to not be tricked by it, with the exception of niche products.

In a second step, we proposed an alternative approach that quantified the evolving credibility of reviews, which could be used to indicate to customers the availability of trustworthy information about products. In a third step, we further implemented our approach by using a publicly available dataset that labelled non-credible and potentially fake reviews.

To our knowledge, this study is one of the first to propose an alternative credibility approach and identify fakery’s economic impact. A major managerial implication is that firms ought to focus on displaying information, thus ensuring that the most informative product attributes, including the most informative reviews, are featured prominently for customers, rather than expending more effort on the elusive task of fakery detection.

6.1. Limitations

This paper has some limitations that we discuss here. The first limitation is that the paper’s findings are based on a single – albeit giant – retailer, Amazon. This retailer had a set of featuring techniques in place that were very specific to this site. These techniques were, however, quite efficient at eliciting additional information from the customer base through the helpful vote. They therefore make a great set of techniques that other retailers may want to use to combat potential fakery. Additional potential limitations are summarised in Table 5, along with the approaches adopted in this paper to address them, albeit partially.

6.2. Future directions

As fakery becomes an increasingly difficult problem to address, this paper proposed an alternative approach that relies on credibility. While we studied the domain of reviews, another domain impacted by fakery is news. It would therefore be quite valuable to investigate the performance of our approach in the context of fake news. The challenges there might, however, be greater, as news – especially political news – evolves quickly, which would require further adaptation of our approach.

Table 5. Potential Limitations and Methods to Address Them.

| Data solely from Amazon.com | This retailer is a major player in online reviews and as such quite representative. |
| Limited Number of Products | Product selection ensured a wide range of variability along product dimensions suggested by the literature. # A large number of reviews posted on those products. |
| Validity of using system features as a proxy for credibility (Verified purchase, Vine program, Helpful vote) | Only extremely restrictive criteria on these features were adopted to ensure validity.# A battery of robustness checks were employed. |
| Amazon’s direct influence on rank through Sponsored Products | Dataset is large enough to consist in its majority of “Non-Sponsored Products” |

Notes

1. Amazon.com features reviews based on helpful votes from the customer base. This featuring varies over time. Since the data collection recurred every other week, we were able to capture what helpful votes had been posted in each time period, what sequence of reviews had been presented during each time period, etc. This information could not have been extracted if the data collection occurred only once.
2. Amazon recently changed this label to “Top Reviews,” although the basis of identifying such reviews did not change.
3. Gamma distribution is traditionally used to measure the time between the occurrence of events when the event process is not completely random (Mun, 2008). Here, for each review s featured, there were $s_i$ helpful votes arriving, with each wait time independently distributed as Exponential(λ).
4. The other option was to use the department within which Amazon listed the product. However, the department-level categorisation was generally too broad to have products with different sales patterns and helpful voting patterns.
5. For additional validity, we conducted a variety of tests to ensure the validity of the Arellano–Bond GMM estimation. The first was a test of overidentifying restrictions that checked whether the instruments were valid, with $H_0$ being that the overidentifying restrictions were valid. This could be accomplished through a Hansen test with a $\chi^2$ distribution. As listed in Table 3, the test failed to reject the null hypothesis, thus ensuring that the models selected were appropriate. The test of overidentification was not rejected at $p = 0.711$ for (1.1), $p = 0.421$ for (1.2), and $p = 0.719$ for (1.3). The second test examined the error term for serial correlation. The test entailed examining whether the differentiated error term was second-order serially correlated – while AR(1) was expected in first
differences. Therefore the test investigated AR(2) in differences to check for AR(1) in levels. As listed in Table 3, the first-order serial correlation for the differed error term was significant with $p < 0.001$, and the second-order serial correlation was not significant, with $p = 0.19$ for (1,1), $p = 0.821$ for (1,2), and $p = 0.176$ for (1,3).

**Disclosure statement**

No potential conflict of interest was reported by the author.

**Funding**

This work was supported by the Social Sciences and Humanities Research Council of Canada [30-2013-000562].

**ORCID**

Wael Jabr  [http://orcid.org/0000-0001-5850-5077](http://orcid.org/0000-0001-5850-5077)

**References**

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*(3), 488–500. [https://doi.org/10.2307/1879431](https://doi.org/10.2307/1879431)

Ananthakrishnan, U. M., Li, B., & Smith, M. D. (2019). A tangled web: Should online review portals display fraudulent reviews? SSRN. [https://ssrn.com/abstract=3297363](https://ssrn.com/abstract=3297363)

Anderson, T. W., & Cheng, H. (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics, 18*(1), pp. 47–82.

Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association, 76*(375), pp. 598–606.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science, 57*(8), 1485–1509. [https://doi.org/10.1287/mnsc.1110.1370](https://doi.org/10.1287/mnsc.1110.1370)

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*(1), 29–51. [https://doi.org/10.1016/0304-4076(94)01642-D](https://doi.org/10.1016/0304-4076(94)01642-D)

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research, 53*(3), 229–297. [https://doi.org/10.1509/jmr.14.0380](https://doi.org/10.1509/jmr.14.0380)

Bakos, Y. (1998). The emerging role of electronic marketplaces on internet. *Communications of the ACM, 41*(8), 35–42. [https://doi.org/10.1145/280324.280330](https://doi.org/10.1145/280324.280330)

Bapna, R., Langer, N., Mehra, A., Gopal, R., & Gupta, A. (2013). Human capital investments and employee performance: An analysis of IT services industry. *Management Science, 59*(3), 641–658. [https://doi.org/10.1287/mnsc.1120.1586](https://doi.org/10.1287/mnsc.1120.1586)

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143. [https://doi.org/10.1016/S0304-4076(98)00009-8](https://doi.org/10.1016/S0304-4076(98)00009-8)

Branco, F., Sun, M., & Villas-Boas, J. M. (2016). Too much information? Information provision and search costs. *Marketing Science, 35*(4), 605–618. [https://doi.org/10.1287/mksc.2015.0959](https://doi.org/10.1287/mksc.2015.0959)

Brooks, R. C. (1957). Word-of-mouth advertising in selling new products. *Journal of Marketing, 22*(2), 154–161. [https://doi.org/10.1177/000222495702200205](https://doi.org/10.1177/000222495702200205)

Chen, H., Zheng, Z., & Ceran, Y. (2016). De-biasing the reporting bias in social media analytics. *Production & Operations Management, 25*(5), 849–865. [https://doi.org/10.1111/poms.12509](https://doi.org/10.1111/poms.12509)

Chen, P.-Y., Dhanasobhon, S., & Smith, M. D. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at Amazon.com. SSRN.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science, 54*(3), 477–491. [https://doi.org/10.1287/mnsc.1070.0810](https://doi.org/10.1287/mnsc.1070.0810)

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research (JMR), 50*(4), 463–476. [https://doi.org/10.1509/jmr.12.0063](https://doi.org/10.1509/jmr.12.0063)

Cheung, M. Y., Chuan, L., Choo Ling, S., & Huaping, C. (2009). Credibility of electronic word-of-mouth: informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce, 13*(4), 9–38. [https://doi.org/10.2753/JIE1086-441310402](https://doi.org/10.2753/JIE1086-441310402)

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science, 37*(5), 688–709. [https://doi.org/10.1287/mksc.2018.1090](https://doi.org/10.1287/mksc.2018.1090)

Chunhua, W., Hai, C., Chan, T. Y., & Xianghua, L. (2015). The economic value of online reviews. *Marketing Science, 34*(5), 739–754. [https://doi.org/10.1287/mksc.2015.0926](https://doi.org/10.1287/mksc.2015.0926)

Cialdini, R. B. (1985). *Influence: Science and practice*. Scott, Foreman.

Competition and Markets Authority. (2015). *Online reviews and endorsements*. U.K. Government. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/436238/Online_reviews_and_endorsements.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/436238/Online_reviews_and_endorsements.pdf)

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407–1424. [https://doi.org/10.1287/mnsc.49.10.1407.17308](https://doi.org/10.1287/mnsc.49.10.1407.17308)

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science, 52*(10), 1577–1593. [https://doi.org/10.1287/mnsc.1060.0567](https://doi.org/10.1287/mnsc.1060.0567)

Dellarocas, C., & Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science, 54*(3), 460–476. [https://doi.org/10.1287/mnsc.1070.0747](https://doi.org/10.1287/mnsc.1070.0747)

Dhanasobhon, S., Chen, P., Smith, M., & Chen, P. (2007). An analysis of the differential impact of reviews and reviewers at Amazon.com. *International Conference on Information Systems, Montreal, QC, Canada.*

Edelman, B. (2017). The market design and policy of online review platforms. *Oxford Review of Economic Policy, 33*(4), 635–649. [https://doi.org/10.1093/orp/grx043](https://doi.org/10.1093/orp/grx043)

European Union. (1984). *Council directive concerning misleading and comparative advertising*. [https://eur-lex.europa.eu/eli/dir/1984/450/2005-06-12](https://eur-lex.europa.eu/eli/dir/1984/450/2005-06-12)

Federal Trade Commission. (2008). *Unfair or deceptive acts or practices*. [https://www.federalreserve.gov/boarddocs/ supmanual/ccc/fcca.pdf](https://www.federalreserve.gov/boarddocs/ supmanual/ccc/fcca.pdf)

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A
Meta-analysis. *Journal of Retailing, 90*(2), 217–232. https://doi.org/10.1016/j.jretai.2014.04.004

Fogg, B., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., Boyd, J., & Brown, B. (2001) What makes Web sites credible? A report on a large qualitative study. Paper presented at the Conference on Human Factors in Computing Systems, Seattle, WA.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research, 19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193

Fradkin, A., Grewal, E., Holtz, D., & Pearson, M. (2018) Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. Paper presented at the Proceedings of the Sixteenth ACM Conference on Economics and Computation.

Glazer, J., Herrera, H., & Perry, M. (2019). Fake persuasion: The market for product reviews. CEPR. https://ssrn.com/abstract=3266438

Godes, D., & Silva, J. C. (2012). Sequential &temporal dynamics of online opinion. *Marketing Science, 31*(3), 448–473. https://doi.org/10.1287/mksc.1110.0653

Godinho De Matos, M., Ferreira, P., Smith, M. D., & Telang, R. (2016). Culling the Herd: Using real-world randomized experiments to measure social bias with known socially goods. *Management Science, 62*(9), 2563–2580. https://doi.org/10.1287/mnsc.2015.2258

Gu, B., Park, J., & Konana, P. (2012). The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research, 23*(1), 182–196. https://doi.org/10.1287/isre.1110.0343

Gunther, A. C. (1992). Biased press or biased public? Attitudes toward media coverage of social groups. *Public Opinion Quarterly, 56*(2), 147–167. https://doi.org/10.1086/269308

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(1), 635–650. https://doi.org/10.1086/266350

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly, 41*(2), 449–4417. https://doi.org/10.25300/MISQ/2017/41.2.06

Jabr, W., & Zheng, Z. (2014). Know yourself and know your enemy: An analysis of firm recommendations and consumer reviews in a competitive environment. *MIS Quarterly, 38*(3), 635–654. https://doi.org/10.25300/MISQ/2014/38.3.01

Jacoby, J., Speller, D. E., & Berning, C. K. (1974). Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research, 1*(1), 31–43. https://doi.org/10.1086/208579

Keye, B. K., & Johnson, T. J. (2011). Hot diggity blog: A cluster analysis examining motivations and other factors for why people judge different types of blogs as credible. *Mass Communication and Society, 14*(2), 236–263. https://doi.org/10.1080/15205431003687280

King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don’t know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing, 28*(3), 167–183. https://doi.org/10.1016/j.intmar.2014.02.001

Korffatis, N., Garcia-Bariocanal, E., & Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications, 11*(3), 205–217. https://doi.org/10.1016/j.elerap.2011.10.003

Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2018). Detecting anomalous online reviewers: An unsupervised approach using mixture models. SSRN. https://ssrn.com/abstract=3258708

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research, 27*(4), 940–961. https://doi.org/10.1287/isre.2016.0674

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lee, S.-Y., Qiu, L., & Whinston, A. (2014) Manipulation: Online platforms' inescapable fate. Paper presented at the Thirty Fifth International Conference on Information Systems, Auckland, New Zealand.

Lee, Y.-J., Hosanagar, K., & Yong, T. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science, 61*(9), 2241–2258. https://doi.org/10.1287/mnsc.2014.2082

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research, 19*(4), 456–474. https://doi.org/10.1287/isre.1070.0154

Li, X., Hitt, L. M., & Zhang, Z. J. (2011). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems, 27*(4), 9–42. https://doi.org/10.2753/MIS0742-122270401

Lu, S. F., & Rui, H. X. (2018). Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Science, 64*(6), 2557–2573. https://doi.org/10.1287/mnsc.2017.2741

Luca, M. (2017). Designing online marketplaces: Trust and reputation mechanisms. *Innovation Policy and the Economy, 17*(1), 77–93. https://doi.org/10.1086/688845

Luca, M., & Zervas, G. (2016). Fake It till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412–3427. https://doi.org/10.1287/mnsc.2015.2304

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review, 104*(8), 2421–2455. https://doi.org/10.1257/aer.104.8.2421

Moe, W. W., & Fader, P. S. (2004). Dynamic conversion behavior at E-commerce sites. *Management Science, 50*(3), 326–335. https://doi.org/10.1287/mnsc.1040.0153

Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research, 48*(3), 444–456. https://doi.org/10.1509/jmkr.48.3.444

Mulherjee, A., Liu, B., & Glance, N. (2012, April). Spotting fake reviewer groups in consumer reviews. 21st International World Wide Web Conference, Lyon, France.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review, 95*(4), 1031–1053. https://doi.org/10.1257/0002828054825619
Mun, J. (2008). Appendix C: Understanding and choosing the right probability distributions. In J. Mun (Ed.), Advanced analytical models: Over 800 models and 300 applications from the base II accord to Wall Street and beyond (Vol. 419, pp. 899–917), John Wiley & Sons.

Nagle, F., & Riedl, C. (2017). Online word of mouth and product review disagreement. Harvard Business School. https://ssrn.com/abstract=2259055

Nagler, M. G. (1993). Rather bait than switch: Deceptive advertising with bounded consumer rationality. *Journal of Public Economics, 51*(3), 359–378. https://doi.org/10.1016/0047-2727(93)90071-Z

Peltzman, S. (1981). The effects of flc advertising regulation. *Journal of Law & Economics, 24*(3), 403–448. https://doi.org/10.1086/66992

Qiu, Y., Gopal, A., & Hann, I.-H. (2017). Logic pluralism in mobile platform ecosystems: A study of indie app developers on the iOS app store. *Information Systems Research, 28*(2), 225–249. https://doi.org/10.1287/isre.2016.0664

Roodman, D. (2006). How to do Xiabond2: An introduction to difference and system GMM in Stata. Center for Global Development. https://ssrn.com/abstract=982943

Saulnier, C. G. (2015). *Can you trust what you read on the internet? Designing platforms to deal with shill reviews*. Job Market Paper.

Schlosser, A. E., White, T. B., & Lloyd, S. M. (2006). Converting web site visitors into buyers: How web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of Marketing, 70*(2), 133–148. https://doi.org/10.1509/jmkg.70.2.133

Self, C. (1996). Credibility. In M. B. Salwen & D. W. Stacks (Eds.), *An integrated approach to communication theory and research*. Lawrence Erlbaum.

Shukla, A., Wang, W., Gao, G. G., & Agarwal, R. (2019). *Catch me if you can — detecting fraudulent online reviews of doctors using deep learning*. SSRN. https://ssrn.com/abstract=3320258

Sun, M. (2012). How does the variance of product ratings matter? *Management Science, 58*(4), 697–707. https://doi.org/10.1287/mnsc.1110.1458

Watson, J., Ghosh, A. P., & Trusov, M. (2018). Swayed by the numbers: The consequences of displaying product review attributes. *Journal of Marketing, 82*(6), 109–131. https://doi.org/10.1177/0022242918045468

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics, 126*(1), 25–51. https://doi.org/10.1016/j.jeconom.2004.02.005

Xiang, Y., & Sarvary, M. (2007). News consumption and media bias. *Marketing Science, 26*(5), 611–628. https://doi.org/10.1287/mksc.1070.0279

Yelp Support Center. (2016). *Why would a review not be recommended?* Retrieved April 20, 2017, from https://www.yelp-support.com/article/Why-would-a-review-not-be-recommended?lang=en_US

Yin, D., Mitra, S., & Zhang, H. (2016). When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Information Systems Research, 27*(1), 131–144. https://doi.org/10.1287/isre.2015.0617

Zervas, G., Proserpio, D., & Byers, J. (2015). A first look at online reputation on airbnb, where every stay is above average. SSRN.

Zhiqiang, Z., Fader, P., & Padmanabhan, B. (2012). From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Information Systems Research, 23*(3), 698–720. https://doi.org/10.1287/isre.2011.0385

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing, 74*(2), 133–148. https://doi.org/10.1509/jm.74.2.133

Zhu, Y., & Dukes, A. (2015). Selective reporting of factual content by commercial media. *Journal of Marketing Research, 52*(1), 56–76. https://doi.org/10.1509/jmr.12.0379

Zinman, J., & Zitzewitz, E. (2016). Wintertime for deceptive advertising? *American Economic Journal, 8*(1), 177–192. http://dx.doi.org/10.1257/app.20130346
Appendix I – Snippets from government and professional literature on review fakery

The Best Book Reviews Money Can Buy

Nick Oxford for The New York Times

Todd Jason Rutherford inside his home in Bixby, Okla. He says that he is now suspicious of all online reviews — whether of books or of anything else.

By DAVID STREITFELD

Published: August 25, 2012 331 Comments

Give Yourself 5 Stars? Online, It Might Cost You

By DAVID STREITFELD

Published: September 22, 2013

“1 celebrate myself, and sing myself,” wrote Walt Whitman, America’s great bard of self-promotion. As the world goes ever more digital, quite a few businesses are adopting that philosophy — hiring a veritable chorus of touts to sing their nonexistent praises and lure in customers.

Enlarge This Image

Nathaniel Brooks for The New York Times

Eric T. Schneiderman, the New York attorney general, called the deceptions “worse than old-fashioned false advertising.”
Fake Reviews: Amazon's Rotten Core

The web has created some fantastic opportunities for authors, publishers and self-publishers alike, but this summer has seen the industry's dark underbelly revealed in all its venal, pustulant ignominy. Things kicked off in July at the [Theakston Old Peculier Crime Writing Festival](http://www.theakston.org.uk/) where successful author Stephen Leather confessed, during an on-stage panel discussion, that he used fake accounts to promote his own books. This admission of sockpuppetry shocked the writing community and has been [covered well](http://www.amazon.com) by fellow panellist Steve Mosby.
Amid Fake Reviews, Consumers Are Skeptical of Social Media Marketing

To paraphrase a famous line, truth is the first casualty of marketing.

With the explosion of online marketing on social media platforms, consumer skepticism about the veracity and independence of online reviews of products and services is running high.

A new survey about attitudes to online marketing techniques, such as fake Facebook 'likes', tweets about brands, and hiding negative reviews from search engine results, suggests that marketers are a lot more relaxed about such practices than the public.

The survey, conducted among 3,000 consumers and 1,000 marketers in the U.K. by pollster YouGov PLC for the Chartered Institute of Marketing, found...
Appendix II – Excerpts from the Competition and Markets Authority of the UK published in 2015

Due to concerns about the potential for reviews and endorsements to mislead consumers and distort their decisions (and negatively impact consumers and businesses), the Competition and Markets Authority (CMA) of the UK launched in February 2015 a call for information, thus inviting all interested parties to submit any relevant information, on online reviews and endorsements, to understand these important sectors better. The CMA has found that online reviews and endorsements are proving useful to consumers. The agency reports that between 76% and 80% of consumers (across the six sectors) who had used reviews before making a purchase thought that it was either ‘very likely’ or ‘fairly likely’ that the reviews they read were written by genuine consumers. Their consumer survey also suggests that goods and services often match up to consumers’ expectations after they use review sites to inform purchases with

- 18 to 26% of consumers found that the product or service was ‘much better’ or ‘a little better’ than the impression the online reviews had given;
- 61 to 70% of consumers found that the product or service was ‘about the same’ as the impression the online reviews had given; and
- only 4 to 8% of respondents felt that their experience was ‘a little worse’ or ‘much worse’ than the impression the online reviews had given.

The CMA had also has concerns that some review posting practices may be unlawful. Their research showed that estimates of the proportion of suspected fake reviews that are published on review sites vary widely. Some estimates though put this number at around 1 to 2% with the acknowledgment that given the clandestine nature of fake reviews, it would be almost impossible to arrive at a credible figure. Furthermore, from the information CMA gathered, it seems that fake negative reviews may be more of a risk for ‘open’ review systems than ‘closed’ ones (in a closed systems only a confirmed buyer of the product or service is able to submit a review. With ‘closed’ systems, the product or service must be purchased before a review can be submitted, and this likely makes it more difficult to leave fake negative reviews in significant numbers.

Appendix III – Dataset Statistics

| Category Type               | Product Count |
|-----------------------------|---------------|
| Automotive                  | 176           |
| Grocery & Gourmet Food      | 290           |
| Health & Personal Care      | 272           |
| Home & Kitchen              | 262           |
The table below displays the “productivity” of reviewers:

| Review Count | Reviewers at this level |
|--------------|-------------------------|
| 2            | 143,728                 |
| 3            | 40,614                  |
| 4            | 15,685                  |
| 5-10         | 10,835                  |
| 11-20        | 278                     |
| 21+          | 6,823                   |

The table below shows the correlation matrix for equation 1.R:

|                  | 1         | 2         | 3         | 4         | 5         | 6         | 7         |
|------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Avg. Recent Rating | 1         | -0.7665   | 1         |           |           |           |           |
| Std. Dev. Recent Rating | 0.0853   | -0.0053   | 0.0261   | 1         |           |           |           |
| Total Reviews    | 0.4043    | 0.0043    | 0.0395   | 1         |           |           |           |
| Featured Recent Reviews | 0.5181   | -0.4565   | -0.0022  | -0.0633   | 1         |           |           |
| Overall Rating   | 0.1137    | -0.1214   | 0.0853   | -0.0311   | 0.2873    | 0.0505    | 1         |
| Ln Price         |           |           |           |           |           |           |           |
| Avg. Help Rating |           |           |           |           |           |           |           |

The table below shows the correlation matrix for equation 1.H:

|                  | 1         | 2         | 3         | 4         | 5         | 6         | 7         |
|------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Avg. Help Rating | 1         | -0.6857   | 1         |           |           |           |           |
| Std. Dev. Help Rating | 0.0853   | -0.0053   | 0.0261   | 1         |           |           |           |
| Featured Help Reviews | -0.2001  | 0.2662    | -0.3352  | 1         |           |           |           |
| Overall Rating   | 0.2873    | -0.2271   | -0.0022  | 0.0378    | 1         |           |           |
| Ln Price         | 0.0505    | -0.0602   | 0.1597   | -0.3812   | -0.021    | 1         |           |
| Avg. Recent Rating | 0.1137   | -0.0955   | -0.0053  | 0.0802    | 0.5181    | -0.0451   | 1         |

Appendix IV – Derivation of Gamma Distribution for String S of Helpful Reviews

If X and Y are independent gamma-distributed random variables (RV) with parameters $s_x, \lambda$ and $s_y, \lambda$ respectively, then the sum of those RVs, $X + Y$, is also a gamma-distributed RV with parameters $s_x + s_y, \lambda$.

The proof is well-established and follows from the calculation

$$P_{X+Y}(z) = (P_X + P_Y)(z)$$

$$P_{X+Y}(z) = \frac{1}{\Gamma(s_x)\Gamma(s_y)} \int_0^z \lambda^{s_x-1}(\lambda z)^{s_x-1} \lambda^{s_y-1}(\lambda z)^{s_y-1} dx$$

where $C$ is a constant.

Since $P_{X+Y}$ is a density, thus integrating to 1, the derivation of $C$ is:

$$C = \int_0^\infty e^{-\lambda z^{s_x+s_y-1}} dz = \frac{\lambda^{s_y}}{\Gamma(s_x+s_y)}$$

Combining G1 and G2 shows that
By induction, it follows that if $X = X_1 + \ldots + X_r + \ldots + X_R$ are independent gamma-distributed random variables with parameters $s_i, \lambda$, then the sum $X = X_1 + \ldots + X_r + \ldots + X_R$ is a gamma-distributed RV with parameters $\left( \sum_{i=1}^{R} s_i, \lambda \right)$.

**Appendix V – GMM Estimation**

The use of instrumentation removes the potential endogeneity bias resulting from the correlation between the regressor and the error term (Wooldridge, 2010). Anderson and Hsiao (1981) and Anderson and Cheng (1982) were the pioneers in proposing use of the GMM procedure within a dynamic context and used differencing in order remove the fixed effects in the error term which are correlated with the lagged dependent variable. The difference of the lagged dependent variable however will still be correlated with the error term and, hence, should be instrumented. These researchers proposed using the second lag of the dependent variable or its lagged difference as instruments since those are expected to be uncorrelated to the error term. Arellano and Bond (1991) and Kiviet (1995) analysed the properties of the two instruments suggested by Anderson and Cheng (1982) and found that the “level” instrument has smaller variance and is, hence, superior to the “differenced” one.

Arellano and Bond (1991) suggested exploiting an enlarged set of instruments; namely, all available lagged values of the dependent variable and the lagged values of the exogenous regressors. Arellano and Bover (1995), and Blundell and Bond (1998) consequently suggested using additional information contained in levels, which should result in more efficient estimator, known as a system-GMM estimator. In the system-GMM estimator, both predetermined and endogenous variables in first differences are instrumented with suitable lags of their own levels (used by Arellano-Bond); and predetermined and endogenous variables in levels are instrumented with suitable lags of their own first differences. As a consequence, the system-GMM estimator should produce more efficient estimates and, hence, outperform the difference-GMM estimator.

A crucial assumption for the validity of GMM is that generated instruments are exogenous, i.e. do not correlate with the error term. Sargan and Hansen-J tests (Hansen, 1982) have been designed to detect violation of this assumption, but there is no formal test to check how many instruments should be cut (Ruud, 2000). Sargan and Hansen-J set the null as “instruments are valid”, which is the assumption investigated. Sargan/Hansen tests can be also used to test the validity of subsets of instrument, through the difference-in-Sargan specification.

There are two additional advantages of the GMM estimator. First, it does not require any distributional assumptions, like normality, which then should be subject of diagnostic testing. Second, it can allow for heteroskedasticity of unknown form by estimating “robust” parameters.

**Appendix VI – Robustness Checks for the Validity of the Analysis**

We conduct two robustness checks vis-à-vis the validity of using Verified Purchase and Vine Voice reviews, as well as Helpful Votes, as tools to conduct the analysis. First, one might argue that fakers could game the system by purchasing the products they are manipulating before they post fake reviews, so that the reviews still end up labeled as verified. This is a plausible argument, which, if correct, would negate the validity of the falsification analysis. However, Table V1 shows that an overwhelming number of products in the sample had more than 50% of their reviews labeled as Verified Purchases (only 124 products have less than 50% of their reviews as verified). So the proposition that fakers might purchase products so that their fake reviews pass as verified becomes an expensive one (literally), given the overwhelming number of verified reviews. Additionally, Table V1 details the prevalence of verified reviews. Column 1 shows the share of verified reviews from the total number of reviews (starting from 10% to more than 90% of reviews being verified). Then for each level of verified reviews, column 2 show the total number of products with that level of verified reviews share.

Another argument is that Vine Voice reviews could also be labeled as fakey. Table V2 shows that, for an overwhelming majority of the products in the sample, just 1% of the reviews were vine and less than 50 products had more than 10% of their reviews originating from vine customers. These findings confer a greater validity to the paper’s conclusions.

Tables V1 and V2 provide greater insights into the functioning of Amazon’s online review forum. The use of the various featuring tools (Verified Purchase, Vine Voice and Helpful Vote) is at levels that would suggest a healthy functioning of the

**Table V1.** Propensity of reviews labeled as Verified.

| Percentage Verified Reviews to Total Reviews | Proportion of products in sample | Proportion of products in sample with price between $50 & $100 | Proportion of products in sample with price greater than 100 |
|---------------------------------------------|---------------------------------|---------------------------------------------------------------|------------------------------------------------------------|
| 10.00%                                      | 0.20%                           | 0.00%                                                        | 0.00%                                                       |
| 25.00%                                      | 1.52%                           | 0.92%                                                        | 2.00%                                                       |
| 50.00%                                      | 15.92%                          | 11.93%                                                       | 10.00%                                                      |
| 60.00%                                      | 20.08%                          | 21.10%                                                       | 24.00%                                                      |
| 70.00%                                      | 31.95%                          | 33.03%                                                       | 22.00%                                                      |
| 80.00%                                      | 23.12%                          | 26.61%                                                       | 32.00%                                                      |
| 90.00%                                      | 6.49%                           | 5.50%                                                        | 10.00%                                                      |
| More                                        | 0.71%                           | 0.92%                                                        | 0.00%                                                       |
orum in vetting individual reviews. While this is based on the sample products being tracked, it is worth mentioning that the levels of verified, vine and help vote were not criteria in the selection of products to track, but rather an outcome.

External Robustness Check: With so much at stake for the validity of the falsification analysis of the fakery algorithms adopted, we conducted further robustness checks of those three fakery algorithms. To that end we used a dataset of restaurant reviews and reviewers from Yelp.com. This online infomediary conducts some curation of their reviews, separating them into recommended and non-recommended reviews. Details of the curation process are not publicly available, but the highlights of the approach are available and rely on the quality of a review, and the reliability and activity of the reviewer as crucial components. Using the same falsification process from earlier, we ran the fakery algorithms on the Yelp dataset of recommended reviews consisting of 2,225,213 reviews, 552,339 reviewers and 77,079 businesses. Using identical credibility criteria, the fakery algorithms ended up misclassifying reviewers as fakers at a rate of 31.33% (or 173,179 out of 552,339 Credible Reviewers when using the 90th percentile distribution of scores obtained from the fakery detection algorithms). This misclassification rate increased to 43.57% when we slightly loosened the fakery detection algorithms threshold to the 75th percentile. Once again, these numbers indicate that widely-used fakery algorithms commit high rates of misclassification and thus should be used with caution to detect fakery in online infomediaries.

| Percentage of Vine Reviews to Total Reviews | Proportion of products in sample | Proportion of products in sample with price between $50 & $100 | Proportion of products in sample with price greater than $100 |
|---------------------------------------------|---------------------------------|---------------------------------------------------------------|-------------------------------------------------------------|
| 1.00%                                       | 86.71%                          | 86.00%                                                        | 81.65%                                                      |
| 2.50%                                       | 5.88%                           | 6.00%                                                         | 10.09%                                                      |
| 5.00%                                       | 3.04%                           | 8.00%                                                         | 4.59%                                                       |
| 10.00%                                      | 2.94%                           | 0.00%                                                         | 2.75%                                                       |
| 25.00%                                      | 1.12%                           | 0.00%                                                         | 0.92%                                                       |
| 50.00%                                      | 0.30%                           | 0.00%                                                         | 0.00%                                                       |
| More                                        | 0.00%                           | 0.00%                                                         | 0.00%                                                       |