# The Market for Fake Reviews

He, Sherry and Hollenbeck, Brett and Proserpio, Davide

UCLA Anderson, UCLA Anderson, USC Marshall

21 January 2021

# The Market for Fake Reviews

Sherry He
Anderson School of Management
UCLA[*]

Brett Hollenbeck
Anderson School of Management
UCLA[†]

Davide Proserpio
Marshall School of Business
USC[‡]

January, 2021[§]

[*]sherry.he.phd@anderson.ucla.edu

[†]brett.hollenbeck@anderson.ucla.edu

[‡]davide.proserpio@marshall.usc.edu

**Abstract**

We study the market for fake product reviews on Amazon.com. These reviews are purchased in large private internet groups on Facebook and other sites. We hand-collect data on these markets to characterize the types of products that buy fake reviews and then collect large amounts of data on the ratings and reviews posted on Amazon for these products, as well as their sales rank, advertising, and pricing behavior. We use this data to assess the costs and benefits of fake reviews to sellers and evaluate the degree to which they harm consumers. The theoretical literature on review fraud shows that there exist conditions when they harm consumers and other conditions where they function as simply another type of advertising. Using detailed data on product outcomes before and after they buy fake reviews we can directly determine if these are low-quality products using fake reviews to deceive and harm consumers or if they are possibly high-quality products who solicit reviews to establish reputations. We find that a wide array of products purchase fake reviews including products with many reviews and high average ratings. Soliciting fake reviews on Facebook leads to a significant increase in average rating and sales rank, but the effect disappears after roughly one month. After firms stop buying fake reviews their average ratings fall significantly and the share of one-star reviews increases significantly, indicating fake reviews are mostly used by low quality products and are deceiving and harming consumers. We also observe that Amazon deletes large numbers of reviews and we document their deletion policy.

# 1 Introduction

Online markets have from their first days struggled to deal with malicious actors. These include consumer scams, piracy, counterfeit products, malware, viruses, and spam. And yet online platforms have become some of the world's largest companies in part by effectively limiting these malicious actors and retaining consumer trust. The economics of these platforms suggest a difficult tradeoff between opening the platform to outside actors such as third party developers and sellers and retaining strict control over access to and use of the platform. Preventing deceptive or fraudulent actions are key to this tradeoff. Third-party participants may have strong incentives to manipulate platforms, such as increasing their visibility in search rankings via fake downloads (Li et al., 2016), increasing revenue via bot-driven advertising impressions (Gordon et al., 2021), manipulating social network influence with fake followers, manipulating auction outcomes, defrauding consumers with false advertising claims (Rao and Wang, 2017; Chiou and Tucker, 2018; Rao, 2018), or manipulating their seller reputation with fake reviews (Mayzlin et al., 2014; Luca and Zervas, 2016).

We study this last form of deception or fraudulent activity: the widespread purchasing of fake product reviews. Fake reviews may be particularly harmful because they not only deceive consumers into purchasing products that may be of low quality, they also erode the long-term trust in the review platforms that is crucial for online markets to flourish (Cabral and Hortacsu, 2010; Einav et al., 2016; Tadelis, 2016). Therefore, if user feedback and product reviews are not trustworthy, in addition to consumers being harmed platform values may suffer as well.

We study the effect of fake reviews on seller outcomes, consumer welfare, and platform value. Despite this practice being unlawful, we document the existence of a large and fast-

moving online market for fake reviews.[1] Specifically, this market features product sellers posting in private online groups to promote their products and solicit willing customers to purchase them and leave positive reviews in exchange for compensation.[2] These groups exist for many online retailers including Walmart and Wayfair but we focus on Amazon because it is the largest and most developed market. We collect data from this market by sending research assistants into these groups to document what products are buying fake reviews and the duration of these promotions. We then carefully track these products' outcomes on Amazon.com including posted reviews, average ratings, prices, and sales rank. This is the first data of this kind, in that it provides direct evidence on both the fake reviews themselves and on detailed firm outcomes from buying fake reviews.

In general, because consumers value trustworthy information and e-commerce platforms value having good reputations, their incentives should be aligned in that they both want to avoid fake reviews. However, this may not always be the case. In particular, platforms may benefit from allowing fake positive reviews if these reviews increase their revenue via higher sales or prices. It may also be the case that fraudulent reviews are not misleading in the aggregate if higher quality firms are more likely to purchase them than lower quality firms. They could be an efficient method for high quality sellers to solve the "cold-start" problem and establish reputations. Indeed, Dellarocas (2006) shows that this is a possible equilibrium outcome. In an extension of the signal-jamming literature on how firms can manipulate strategic variables to distort beliefs, he shows that fake reviews are mainly purchased by high quality sellers and therefore increase market information under the condition

---

[1]The FTC has brought cases against firms alleged to have posted fake reviews, including a case against a weight-loss supplement firm buying fake reviews on Amazon in February 2019. See: `https://www.ftc.gov/news-events/press-releases/2019/02/ftc-brings-first-case-challenging-fake-paid-reviews-independent`.
On May 22, 2020, towards the end of our data collection window, the UK Competition and Markets Authority (CMA) announced it was opening an investigation into these practices. See: `https://www.gov.uk/government/news/cma-investigates-misleading-online-reviews`

[2]The requirement that these reviews be positive to receive payment and the lack of disclosure are what differentiate "fake" reviews from the more innocuous practice of "incentivized" reviews. In the latter case, sellers offer discounted or free products to potential reviewers in exchange for posting an informative review. While incentivized reviews have raised concerns as well since they may be biased upward, in principle, they allow for authentic feedback and typically involve disclosure.

that demand increases convexly with respect to user rating. Due to the way ratings influence product rankings in search results in competitive markets, it is plausible that this property may hold. Other attempts to model fake reviews have also concluded these may benefit consumers and markets (Glazer et al., 2020; Yasui, 2020). The mechanism is different but this outcome is similar to signaling models of advertising for experience goods. Nelson (1970) and later Milgrom and Roberts (1986) show that separating equilibria exist where higher quality firms are more likely to advertise because the returns from doing so are higher for them. This is because they expect repeat business or word-of-mouth once consumers have discovered their true quality. Both Wu and Geylani (2020) and Rhodes and Wilson (2018) study models of deceptive advertising and conclude that this practice can benefit consumers under the right conditions. To the extent that fake reviews generate higher sales which generate future organic ratings, a similar dynamic may play out in our setting. In this case, fake reviews may be seen as harmless substitutes for advertising rather than as being malicious. It is therefore ultimately an empirical question whether firms and regulators should view fake reviews as representing a significant threat to consumer welfare.

Our research objective is to answer a set of currently unsettled questions about online rating manipulation. How does this market work, in particular what are the costs and benefits to sellers from buying fake reviews? What types of products buy fake reviews and how effective are they? Are consumers ultimately harmed by fake reviews or are they mainly used by high quality products, that is, should they be seen more like advertising or outright fraud? Do fake reviews lead to a self-sustaining increase in sales and organic ratings? These questions can be directly tested using the unique panel nature of our data. In particular, if products purchasing fake reviews continue to receive high ratings from consumers after they stop purchasing reviews and continue to enjoy high sales, it suggests the fake reviews are more akin to advertising and fake reviews are mainly bought by high quality products, potentially to solve a cold-start problem. In this case, consumers may not be harmed by fake reviews and the platform might not want to regulate them too strictly. If, by contrast, once

3

products stop buying fake reviews their ratings fall and they begin to receive many one-star ratings, it suggests that these consumers have been deceived into buying products whose true quality was lower than they expected at the time of purchase and therefore overpaid or missed out on a higher quality alternative. In this case, the consumer harm is clear and direct, and the platform reputation will be harmed as well.

We first employ research assistants to construct a random sample of approximately $1,500$ products observed soliciting fake reviews over a 9-month period. We might expect these products to be new products with very few reviews, or else low quality products with very low ratings from organic reviews that must be countered with fake positive reviews. Instead, we find a wide assortment of product types in many categories, including many products with a very large number of reviews at the time we first observe them buying fake reviews. These products also tend not to have especially low ratings, with an average rating slightly higher than comparable products. Almost none of the sellers purchasing reviews in these markets are well known brands, consistent with research showing that online reviews are more effective and more important for small independent firms compared to brand name firms (Hollenbeck, 2018).

We then track the outcomes of these products before and after the buying of fake reviews using data collected from Amazon. In the weeks after they purchase fake reviews the number of reviews posted increases substantially. Their average rating and share of five-star reviews also increase substantially. Ratings increase by .08 stars on average, and the average number of reviews posted per week increases by 7, roughly doubling the number of reviews they receive compared to before soliciting fake reviews. We also observe a substantial increase in search position and sales rank at this time. The increase in average ratings is short-lived, with ratings falling back to the previous level within 2 to 4 weeks, but the increase in the weekly number of reviews, sales rank, and position in search listings remain substantially higher more than four weeks later.

We also track the long-term outcomes associated with the buying of fake reviews. We

find that the evidence primarily supports the consumer harm view. We track outcomes after the last observed post soliciting fake reviews and find that ratings tend to fall as soon as the seller stops buying fake reviews. The share of reviews that are one star increases substantially at this point as well. This pattern also holds for new products and those with few reviews and is, in fact, stronger for them, suggesting that even for this type of product fake reviews are associated with consumer harm. Text analysis confirms that these one-star reviews are distinctive from one-star reviews posted on comparison products, with a greater focus on product quality.

Finally, we document some facts regarding how the platform regulates fake reviews. We see that a very large share of reviews are ultimately deleted by Amazon. For the products in our data observed buying fake reviews, roughly one-third of their reviews are eventually deleted. The reviews that are deleted are longer than non-deleted reviews, more likely to contain photographs, and more likely to be five stars. The bulk of deleted reviews are those that are posted within one to two months of the fake review solicitation that we observe, but they are deleted with an average lag of over 100 days, thus allowing the short term boost in average ratings and number of reviews that we document. Altogether our results suggest that while Amazon's review deletion policy should reduce the long-term harm to consumers from fake reviews, it is inadequate because there is enough of a short-term boost in sales and ratings that firms find it advantageous to participate in this market, and there is a clear consumer harm as shown in the subsequent increase in one-star reviews.

Prior studies of fake reviews include Mayzlin et al. (2014), who argue that in the hotel industry, independent hotels with single-unit owners have a higher net gain from manipulating reviews. They then compare the distribution of reviews for these hotels on Expedia and TripAdvisor and find evidence consistent with review manipulation. Luca and Zervas (2016) uses Yelp's review filtering algorithm as a proxy for fake reviews, and finds that these reviews are more common on pages for firms with low ratings, independent restaurants, and restaurants with more close competitors. Anderson and Simester (2014) show examples of a

different type of fake review: customers rating apparel products on a brand site who never purchased those products. Ananthakrishnan et al. (2020) show using lab experiments that a policy of flagging fake reviews but leaving them posted can increase consumer trust in a platform.

We contribute to this literature in two primary ways. First, we document the actual market where fake reviews are purchased and characterize the sellers participating in this market. This data gives us a direct look at fake reviews, rather than merely inferring their existence. Second, we observe firm outcomes both before and after they purchase fake reviews. This allows us to understand the short and long term effectiveness of fake reviews and assess whether and when consumers are harmed by them.

This research also contributes to the broader academic study of online reviews and reputation. By now, it is well understood that online reviews affect firm outcomes and improve the functioning of online markets (see Tadelis (2016) for a review.) There is also a growing body of research showing that firms take actions to respond to online reviews, including by leaving responses directly on review sites (Proserpio and Zervas, 2016) and changing their advertising strategy (Hollenbeck et al., 2019). There has always existed a difficult tension in the broader literature on online reviews, coming from the fact that sellers may manipulate or fake their reviews. By documenting the types of sellers purchasing fake reviews and the size and timing of their effects on ratings and reviews, we provide guidance to future researchers on how to determine whether review manipulation is likely in their setting.

Finally, we contribute to the literature on fraudulent activity in marketing. This research studies practices such as fake news on social media (Chiou and Tucker, 2018), and deceptive online advertising (Rao, 2018; Wu and Geylani, 2020). The theoretical literature on deceptive practices has emphasized that there are generally conditions when these might make markets more efficient and possibly even benefit consumers (Dellarocas, 2006; Rhodes and Wilson, 2018). It is therefore up to empirical researchers to document the use of fraudulent practices to inform the debate on how regulators and firms should respond to these practices.

6

The rest of the paper proceeds as follows: Section 2 describes our data collection procedure and the settings of our paper; Section 3 present a discussion of the costs and benefits of buying fake reviews; Section 4 documents the short term changes in outcomes like average ratings, number of reviews, and sales rank in the weeks following the buying of fake reviews; Section 5 documents what happens to these outcomes after sellers stop buying fake reviews, Section 6 discusses the Amazon response to the problem of fake reviews; and, finally, Section 7 discusses our findings and provides concluding remarks.

# 2    Data and Settings

In this section, we document the existence and nature of online markets for fake reviews, and discuss in detail the data collection process and the data we obtained to study fake reviews and their effect on seller outcomes, consumer welfare, and platform value. Specifically, we collected data mainly from two different sources, Facebook and Amazon. From Facebook, we obtained data about sellers and products buying fake reviews, while from Amazon we collect product information such as reviews, ratings, and sales data.

## 2.1    Facebook Groups and Data

Facebook is one of the major platforms that Amazon sellers use to recruit fake reviewers. To do so, Amazon sellers create Facebook private groups where they promote their products by soliciting users to purchase their products and leave a five-star review in exchange for a full refund (and in some cases an additional payment). Discovering these groups is straightforward for interested reviewers; it only requires using the Facebook search engine to retrieve a list of them by searching for "Amazon Review". We begin by documenting the nature of these markets and then describe how we collect product information from them.

**Discovering groups**    We collected detailed data on the extent of Facebook group activity during a 4-month period, from Mar 28, 2019 to July 12, 2019. Each day, we collected the

Electronic copy available at: https://ssrn.com/abstract=3664992

Facebook group statistics for the top-30 groups by search rank, only including groups where sellers recruit fake reviewers. During this period, on average, we identify about 23 fake review related groups every day. These groups are large and quite active, with each having about 16,000 members on average, and 568 fake review requests posted per day per group. We observe that Facebook periodically deletes these groups but that they quickly reemerge.

Within these Facebook groups, sellers can obtain a five-star review that looks organic. Figure 1 shows examples of Facebook posts aimed at recruiting reviewers. Usually, these posts contain words such as "need reviews", "refund after pp (Paypal)" with product pictures. To avoid being detected by Amazon's algorithm, sellers do not directly give reviewers the product link; instead, sellers ask reviewers to search for specific keywords associated with the product and then find it using the title of the product, the product photo, or a combination of the two.

The vast majority of sellers buying fake reviews compensate the reviewer by refunding the cost of the product via a PayPal transaction after the five-star review has been posted (most sellers advertise that they also cover the cost of the PayPal fee and sales tax). Moreover, we observe that roughly 15% of products also offer a commission on top of refunding the cost of the product. The average commission value is $6.24 with the highest observed commission for a review being $15. Therefore, the vast majority of the cost of buying fake reviews is the cost of the product itself.

Reviewers are compensated for creating realistic seeming five-star reviews unlike reviews posted by bots or cheap foreign workers with limited English skills, which are more likely to be filtered by Amazon's fraud detection algorithms. First, the fact that the reviewer buys the product means that the Amazon review is listed as a "Verified Purchase" review; second, reviewers are encouraged to leave lengthy, detailed reviews including photos and videos to mimic authentic and organic reviews.[3] Finally, sellers recruit only reviewers located in the United States, with an `amazon.com` account, and with a history of past good reviews.

---

[3]The fact that these fake reviews are from verified purchases indicates that an identification strategy like the one used in Mayzlin et al. (2014) will not work in settings like ours.

This process differs from "incentivized reviews", where sellers offer free or discounted products or discounts on future products in exchange for reviews. Several features distinguish fake reviews from incentivized reviews. The payment for incentivized reviews is not conditional on the review being positive, whereas reimbursement for fake reviews requires a five-star rating. Incentivized reviews, in principle, contain informative content for consumers, whereas in many cases the reviewer posting a fake review has not used or even opened the product. Finally, incentivized reviews typically involve disclosure in the form of a disclaimer contained in the review itself that the product was received for free or at a discount in exchange for the review. Amazon has at times allowed incentivized reviews and even has formally sponsored them through its Vine program and through its "Early Reviewer Program," but the company considers fake reviews a violation of its terms of service by both sellers and reviewers, leaving them subject to being banned from the platform if caught.

**Discovering products** To discover products that are promoted we rely on research assistants. We assign a few active Facebook groups to each one of them and ask them to select Facebook posts randomly. Given a Facebook post, the goal of the research assistants is to retrieve the Amazon URL of the product. To do so, they use the keywords provided by the seller. For example, in Figure 1, the search words would be "shower self", "toilet paper holder" and "cordless vacuum".

After a research assistant successfully identified the product, we ask them to document the following variables: search keywords, product ID, product subcategory (from the Amazon product page), date of the Facebook post, the earliest post date from the same seller for the same product (if older posts promoting the same product exist), and the Facebook group name.

We use the earliest Facebook post date as a proxy for when the seller began to recruit fake reviewers. To identify when a seller stops recruiting fake reviews for a product we continuously monitor each group and record any new posts regarding the same product by
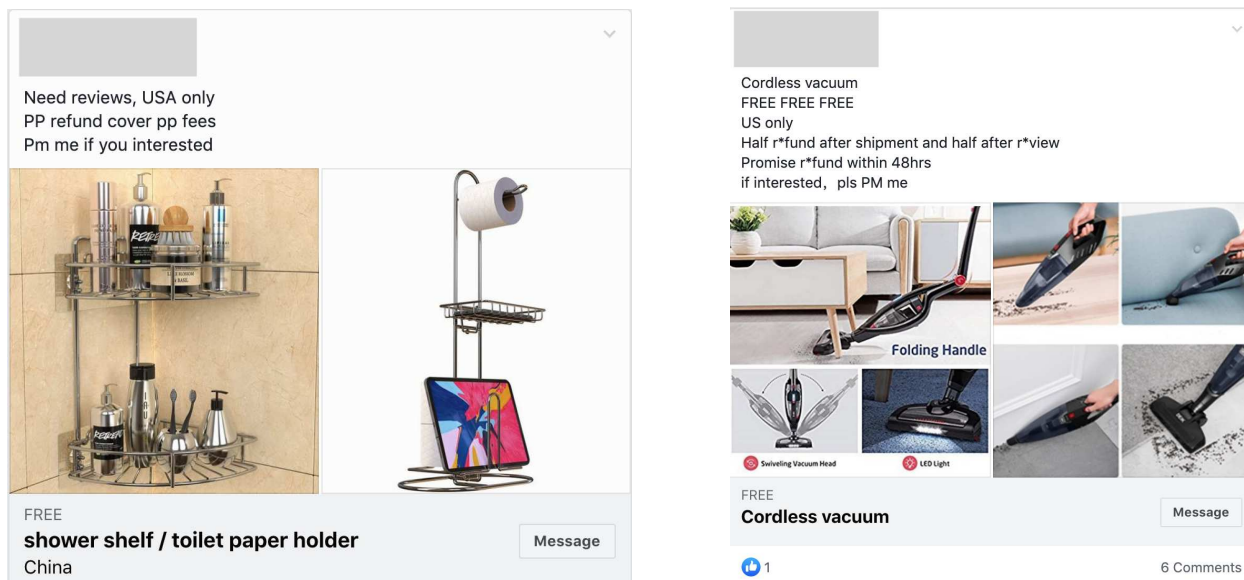
9

Figure 1: Examples of Fake Review Recruiting Posts

searching for the seller's Facebook name or the product keywords. We then use the date of the last observed post as a proxy for when the seller stopped recruiting fake reviews.

We collect data from these Facebook fake review groups using this procedure on a weekly basis from October 2019 to June 2020, and the result is a sample of roughly 1,500 unique products. This provides us with the rough start and end dates of when fake reviews are solicited in addition to the product information.

## 2.2  Amazon Data

We use the Amazon information obtained by the research assistants to collect data about products buying fake reviews from Amazon.com.

**Search Results Data**   For each product buying fake reviews we collect information from the keyword search page results, i.e., products returned as a result of keyword search query. This set of products is useful to form a competitor set for each product buying fake reviews. We collect this information daily and store all consumer-facing information available on these pages including price, coupon, badges (best seller/ amazon's choice), displayed rating,

10

number of reviews, search page number, shipping information, whether the product buy's sponsored listings, and the product position in each page.[4]

**Review Data**   We collect the reviews and ratings for each of the products observed buying fake reviews on a daily basis. For each review we store the following variables: review data, ratings, product ID, review text, helpful votes.

Additionally, we collect the full set of reviews for each product on a bi-monthly basis. The reason for this is that it allows us to measure to what extent Amazon respond to sellers recruiting reviews by deleting reviews that it deems as potentially fake.

**Sales Rank Data**   We rely on Keepa.com and its API to collect sales rank data for the products soliciting fake reviews. We collect this data twice a week for focal products and any products that appear in the category data discussed above. Amazon reports a measure called Best Seller Rank, whose exact formula is a trade secret, but which translates actual sales within a specific period of time into a ranking of products by sales levels.

## 2.3   Descriptive Statistics

In this subsection we provide descriptive statistics on the set of roughly 1,500 products collected between October 2019 to June 2020.

We use this sample of products to characterize the types of products that purchase fake reviews. On the one hand, we might expect these products to be primarily new products with few or no reviews who are trying to jumpstart sales by establishing an online reputation. On the other hand, these might be products with many reviews and low average ratings, whose sellers resort to fake reviews to improve the product reputation and therefore increase sales.

---

[4]Using page number and product position, we can compute the keyword search position of every product.

Table 1: Focal Product Categories and Subcategories

| Category | N | Subcategory | N |
|---|---|---|---|
| Beauty & Personal Care | 193 | Humidifiers | 17 |
| Health & Household | 159 | Teeth Whitening Products | 15 |
| Home & Kitchen | 148 | Power Dental Flossers | 14 |
| Tools & Home Improvement | 120 | Sleep Sound Machines | 12 |
| Kitchen & Dining | 112 | Men's Rotary Shavers | 11 |
| Cell Phones & Accessories | 81 | Vacuum Sealers | 11 |
| Sports & Outdoors | 77 | Bug Zappers | 10 |
| Pet Supplies | 62 | Electric Back Massagers | 10 |
| Toys & Games | 61 | Cell Phone Replacement Batteries | 9 |
| Patio, Lawn & Garden | 59 | Light Hair Removal Devices | 9 |
| Electronics | 57 | Outdoor String Lights | 9 |
| Baby | 42 | Cell Phone Charging Stations | 8 |
| Office Products | 30 | Electric Foot Massagers | 8 |
| Automotive | 29 | Meat Thermometers & Timers | 8 |
| Arts, Crafts, & Sewing | 21 | Aromatherapy Diffusers | 7 |
| Camera & Photo | 19 | Blemish & Blackhead Removal Tools | 7 |
| Clothing, Shoes & Jewelry | 14 | Cell Phone Basic Cases | 7 |
| Computers & Accessories | 12 | Portable Bluetooth Speakers | 7 |

Table 1 shows a breakdown of the top 20 categories and subcategories for our sample of products. The use of fake reviews is widespread across products and product categories. The top categories are "Beauty & Personal Care", "Health & Household", and "Home & Kitchen", but the full sample of products comes from a wide array of categories as the most represented category still only accounts for just 13% of products, and the most represented product in our sample, Humidifiers, only accounts for roughly 1% of products.

We observe substantial variation in the length of the Facebook fake reviews recruiting period, with some products being promoted for a single day and others promoted for over a

month. The average length of the Facebook promotion period is 23 days and the median is 6 days.

Turning to the product age (measured using the first date the product was listed on Amazon), we find that the mean and median product age when they first begin soliciting fake reviews is 229 days and 156 days, respectively. This suggests that products collecting fake reviews are rarely new and without any reputation. Indeed, out of the 1,500 products we observe, only 17 of them solicit fake reviews in their first week after the product appears on Amazon, and only 94 solicit fake reviews in their first month.

Next, we compare the characteristics of our focal products to a set of competitor products. We define competitor products as those products that appear on the same page of search results for the same product keywords as our focal products. Even with these restrictions, we obtain a set of about 200,000 competitor products.

Table 2 compares the focal products with their competitors over several characteristics. We observe that while they are not extremely new when soliciting fake reviews, the focal products are significantly younger than competitor products, with a median age of roughly 5 months compared to 15 months for products not observed buying fake reviews. Moreover, our focal products charge slightly lower average prices than their competitors, with a mean price of $33 compared to $45 for other products. However, this result is mainly driven by the right tail of the price distribution. Fake review products actually charge a higher median price than their competitors but there are far fewer high priced products among the fake review products than among competitors. This may reflect the fact that a primary cost of buying fake reviews is compensating the reviewer for the price of the product. In other words, the more expensive a product is, the more it costs to buy fake reviews.[5]

---

[5]We show this using a simple model in Section 3

13

Table 2: Comparison of Focal Products to Comparable Products

| | Count | Mean | SD | 25% | 50% | 75% |
|---|---|---|---|---|---|---|
| *Avg Rating* | | | | | | |
|   Fake Review Products | 1,315 | 4.4 | 0.5 | 4.1 | 4.5 | 4.8 |
|   All Products | 203,480 | 4.2 | 0.6 | 4.0 | 4.3 | 4.6 |
| *Number of Reviews* | | | | | | |
|   Fake Review Products | 1,425 | 183.1 | 493.5 | 10.0 | 45.0 | 167.0 |
|   All Products | 203,485 | 451.4 | 2,619.0 | 13.0 | 59.0 | 250.0 |
| *Price* | | | | | | |
|   Fake Review Products | 1,425 | 33.4 | 45.0 | 16.0 | 24.0 | 35.0 |
|   All Products | 236,542 | 44.7 | 154.8 | 13.0 | 21.0 | 40.0 |
| *Sponsored* | | | | | | |
|   Fake Review Products | 1,425 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 |
|   All Products | 236,542 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 |
| *Keyword Position* | | | | | | |
|   Fake Review Products | 1,425 | 21.4 | 16.1 | 8.0 | 16.0 | 33.0 |
|   All Products | 236,542 | 28.2 | 17.3 | 13.0 | 23.0 | 43.0 |
| *Age (days)* | | | | | | |
|   Fake Review Products | 1,305 | 229.8 | 251.1 | 77.0 | 156.0 | 291.0 |
|   All Products | 153,625 | 757.8 | 797.1 | 257.0 | 466.0 | 994.0 |
| *Sales Rank* | | | | | | |
|   Fake Review Products | 1,300 | 73,292.3 | 151,236.4 | 7,893.3 | 26,200.5 | 74,801.5 |
|   All Products | 5,647 | 89,926.1 | 323,028.9 | 5,495.0 | 21,610.0 | 72,563.5 |

Turning to ratings, we observe that products purchasing fake reviews have, at the time of their first observed post, relatively high product ratings. The mean rating is 4.4 stars and the median is 4.5 stars, which are both higher than the average ratings of competitor products. Although, we note that ratings may of course be influenced by previous unobserved

Facebook campaigns. Only 14% of products have initial ratings below four stars and only 1.2% have ratings below three stars, compared to 19.5% and 3% for competitor products. Thus, it appears that products purchasing fake reviews do not seem to do so because they have a bad reputation.

We also examine the number of reviews. The mean number of reviews for focal products is 183, which is driven by a long right tail of products with more than 1,000 reviews. The median number of reviews is 45 and roughly 8% of products have zero reviews at the time they are first seen soliciting fake reviews. These numbers are relatively low when compared to the set of competitor products which has a median of 59 reviews and a mean of 451 reviews. Despite these differences, it seems that most of the focal products are not buying fake reviews because they have very few or no reviews.

The last comparison is in terms of sales. We observe that the focal products have slightly lower sales than competitor products as measured by their sales rank, but the difference is relatively minor.

Turning to brand names, we find that almost none of the sellers in these markets are well-known brands. Brand name sellers may use other channels or avoid buying fake reviews altogether to avoid damage to their reputation. This result is also consistent with research showing that online reviews have larger effects for small independent firms relative to firms with well-known brands (Hollenbeck, 2018).

Finally, to better understand which type of sellers are buying fake reviews we collect one additional piece of information. We take the sellers' name from Amazon and check the U.S. Trademark Office for records on each seller. We find a match for roughly 70% of products. Of these products, the vast majority, 84%, are located in China, more precisely in Shenzhen or Guangzhou in the Guangdong province, an area associated with manufacturing and exporting.

To summarize, we observe purchases of fake reviews from a wide array of products across many categories. These products are slightly younger than their competitors but only a

small share of them are truly new products. They also have relatively high ratings, a large number of reviews, and similar prices to their competitors.

# 3   The Simple Economics of Fake Reviews

We build on the results from the previous section on how the market of fake reviews functions, and briefly show the costs and benefits of buying fake reviews. We start by focusing on the costs to the seller of buying a fake review.

First, to buy 1 fake review, a seller must pay to the reviewer:

$$P(1 + \tau + F_{PP}) + Commission \tag{1}$$

Where $P$ is the product's list price, $\tau$ is the sales tax rate, $F_{PP}$ is the PayPal fee, and Commission refers to the additional cash offered by the seller, which is often zero but is sometimes in the \$5-10 range. After the reviewer buys the product, the seller receives a payment from Amazon of:

$$P(1 - c)$$

Where c is Amazon's commission on each sale. So the difference in payments or net financial cost of 1 review is:

$$P(1 + \tau + F_{PP}) + Commission - P(1 - c) = P(\tau + F_{PP} + c) + Commission$$

This is the share of the list price that is lost to PayPal, Amazon, and taxes, along with the potential cash payment. Along with this financial cost the seller bears the production cost of the product (MC), making the full cost of 1 fake review:

$$Cost = MC + P(\tau + F_{PP} + c) + Commission \tag{2}$$

If we define the gross margins rate as $\lambda$ such that $\lambda = \frac{P-MC}{P}$, we can show that equation 2 becomes

$$Cost = P(1 - \lambda + \tau + F_{PP} + c) + Commission \qquad (3)$$

This defines the marginal cost of a fake review to the seller. The benefit of receiving 1 fake review is a function of how many organic sales it creates $Q_o$ and the profit on those sales, which is:

$$Benefit = Q_o P(\lambda - c) \qquad (4)$$

where again $c$ refers to Amazon's commission from the sale. Setting equations 3 and 4 equal allows us to calculate the breakeven number of organic sales $Q_o^{BE}$. This is the number of extra incremental sales necessary to exactly justify buying 1 fake review. If the seller does not offer an additional cash commission, and the vast majority of sellers do not, this can be written as:

$$Q_o^{BE} = \frac{1 - \lambda + \tau + F_{PP} + c}{\lambda - c} \qquad (5)$$

Where the direct effect of price drops out and this is just a function of the product markup and observable features of the market. We take these market features as known:

- $\tau = .0656$[6]

- $F_{PP} = 2.9\%$

- Amazon commission $c$ varies by category but is either 8% or 15% in almost all cases.[7]

The result for products in the 8% commission categories is:

$$Q_o^{BE} = \frac{1.175 - \lambda}{\lambda - .08} \qquad (6)$$

Thus the breakeven level of incremental sales needed to justify buying 1 fake review is a

---

[6]https://taxfoundation.org/2020-sales-taxes/, we aggregate by taking an average of state and local sales taxes.

[7]https://sellercentral.amazon.com/gp/help/external/200336920

17

simple expression of a product's price-cost margin. It is clear that products with larger markups require fewer incremental organic sales in order to justify a fake review purchase. This is for two reasons that this analysis makes clear. First, because the cost of a fake review is lower since conditional on price the marginal cost is lower, and second because the benefit of an organic sale is larger for products with larger markups.

Figure 2 plots equation 6 where the X-axis is $\lambda$ and the Y-axis is $Q_o^{BE}$. It shows that for products with relatively low markups the breakeven number of organic sales approaches 10 but for products with relatively high markups this number is below 1.

Note that this is not a theoretical model of the full costs and benefits of fake reviews, many of which are not accounted for, including the risk of punishment and the extent to which $Q_o$ varies as a result of product quality. This is merely a simple description of the direct financial costs and benefits sellers face and how they determine the profitability cutoff for $Q_o$. Nevertheless, several direct implications follow from this analysis. First, the economics of fake reviews can be quite favorable for sellers since a fairly small number of organic sales are needed to justify their cost. In practice, cheap Chinese imported products often have very large markups such that these sellers only need to generate roughly 1 additional organic sale to profit from a fake review purchase.

Figure 2: Organic Sales Needed to Justify 1 Fake Review



18
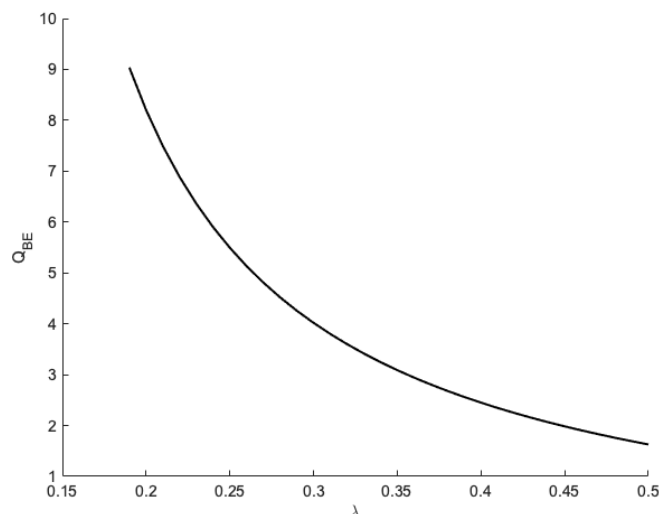
Second, this is especially the case for lower quality products with larger markups. To take a concrete example, imagine two products that both list a price of $25. Product A costs $15 to produce and product B costs $20 to produce because A is of lower quality than B. For product A $Q_o^{BE} = 2.4$ and for product B $Q_o^{BE} = 8.1$. The lower cost product needs far fewer organic sales to justify the expense of 1 fake review.

Third, this analysis makes clear why we are unlikely to observe fake negative reviews applied to competitor products, as in Luca and Zervas (2016) and Mayzlin et al. (2014). The cost of a fake review for a competitor product is significantly higher because it requires the firm buying the review to incur the full price of the competitor's product and the benefit is likely to be lower because the negative effect on competitor sales is indirect and dispersed across potentially many other products.

# 4  Short-term Outcomes from Buying Fake Reviews

In this section, we describe the short-term effects associated with buying fake reviews for important outcomes such as ratings, reviews, and sales rank, as well as other marketing activities such as advertising and promotions. One of the unique features of our data is this detailed panel on firm outcomes observed both before and after sellers buy fake reviews. We therefore quantify the extent to which this practice affects ratings, reviews and sales, although we stress that these results are descriptive in nature. We do not observe the counterfactual outcomes in which these sellers do not buy fake reviews and so the outcomes we measure are not to be interpreted strictly as causal effects.

To evaluate these outcomes, we partition the time around the earliest Facebook recruiting post date (day 0) in 7-day intervals. For example, the interval 0 includes the days in the range [0,7) and the interval -1 includes the days in the range [-7,0). We then plot the quantity of interest for eight 7-day intervals before fake reviews recruiting start and four 7-day intervals after fake reviews recruiting starts. We focus on roughly four weeks after

19

fake reviews recruiting starts because in this section we are interested in discussing short-term effects (recall that the mean length in days of a Facebook campaign is 23 days in our dataset).
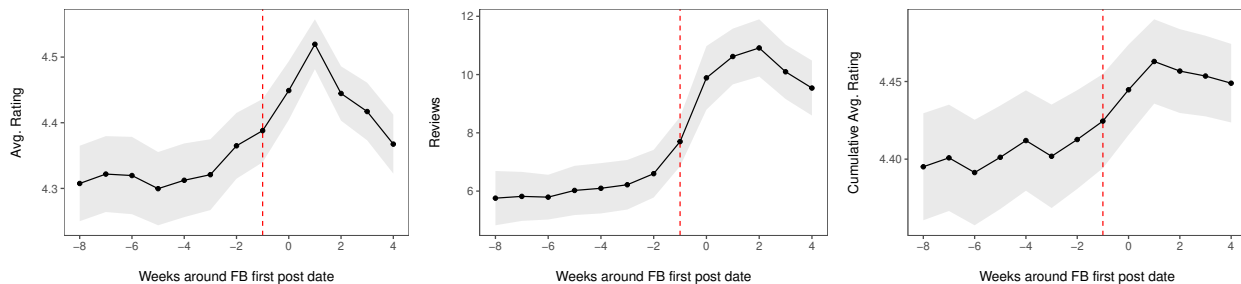


Figure 3: 7-day average ratings, 7-day average number of reviews, and cumulative average ratings before and after fake reviews recruiting begins. The red dashed line indicates the first time we observe Facebook fake review recruiting.

**Ratings and reviews** We start by looking at how ratings and reviews change after the seller begins buying fake reviews. In the left panel of Figure 3 we plot the weekly average rating. Several interesting facts emerge from this figure. First, the average ratings increase by about 5%, from 4.3 stars to 4.5 stars at its peak, after Amazon sellers start recruiting fake reviewers. Second, this positive effect is short and it starts dissipating just two weeks after the beginning of the recruiting of fake reviews; despite this, even after four weeks after the beginning of the promotion, average ratings are still slightly higher than ratings in the pre-promotion period. Third, the average star-rating starts increasing roughly two weeks before the first Facebook post we observe, suggesting that we may not be able to capture with high precision the exact date at which sellers started promoting their products on Facebook. Despite this limitation, our data seems to capture the effect of recruiting fake reviewers fairly well.

Next, we turn to the number of reviews. In the middle panel of Figure 3, we plot the weekly average number of posted reviews. We observe that the number of reviews increases substantially around interval 0, nearly doubling, and confirming the expectation that recruiting fake reviewers is effective at generating new product reviews at a fast pace.

Moreover, and differently from the average rating plot, the effect of recruiting fake reviewers on the number of reviews persist during the entirety of the post-promotion period. Finally, Figure 3 confirms that we are not able to capture the exact data at which the Facebook promotion started.

Does this increase in positive reviews lead to higher displayed product ratings? To answer this question, in the right panel of Figure 3, we plot the cumulative average rating before and after the Facebook promotion starts. We observe a positive change centered around the beginning of the promotion and that stabilized for about two weeks after the promotion begins, after which the effect starts to dissipate.

Finally, we investigate how the effect of recruiting fake reviewers changes with the length of the campaign duration. As discussed in Section 2, there is substantial variation in the length of the Facebook promotion across the products in our dataset. We therefore plot the average rating, reviews, and cumulative average rating by Facebook campaign duration quartiles in Figure 4.[8] Looking at average ratings (left panel), we observe that products with lower ratings in the pre-promotion period seem to be promoted for longer periods than products with higher ratings. Further, and as we would expect, the positive effect of fake reviews is much shorter for products that are promoted for shorter periods. Turning to the number of reviews (middle panel), we observe that products with fewer reviews are promoted for shorted periods; we argue that this might be related to the fact that changing the average ratings of products with a lot of reviews would require collecting more reviews (and therefore more time) than for product with a few reviews. Finally, the cumulative average rating plot (right panel) show patterns similar to those observed for the average ratings; however, the Facebook promotion effect last for a longer period for all products.

**Sales rank**  In the left panel of Figure 5 we plot the average log of sales rank. This figure reveals several facts. First, the figure shows that the sales rank of products that

---

[8]The four quartiles refer to Facebook campaign duration of 1 day, 2-6 days, 7-31 days and more than 31 days.
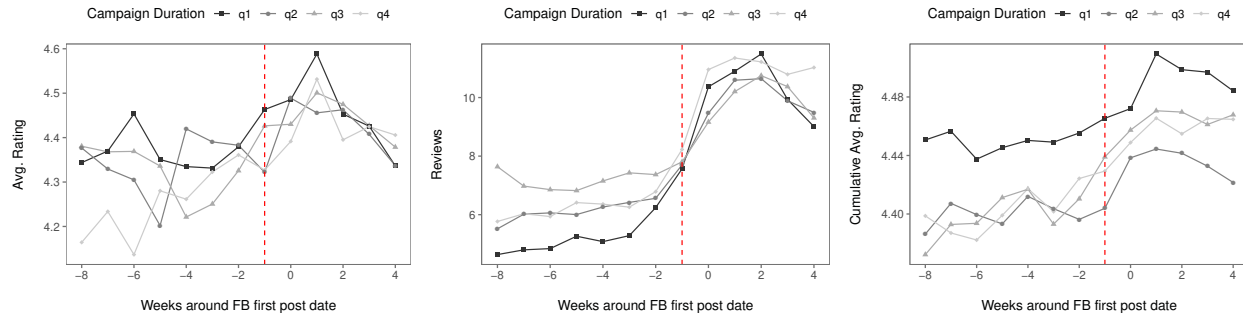
Figure 4: 7-day average ratings, 7-day average number of reviews, and umulative average ratings before and after fake reviews recruiting begins, by Facebook campaign duration quartiles. The red dashed line indicates the first time we observe Facebook fake review recruiting.

are eventually promoted is increasing between the intervals -8 and -3. This suggests that Amazon sellers tend to promote products for which sales are falling. Second, the effect of recruiting fake reviewers is negative (i.e. the sales rank decrease in magnitude and therefore the product sales increase) and large (in the post-promotion period sales rank goes back to the lowest levels observed in the pre-promotion period). Finally, the effect lasts almost all of the post-promotion period.

As we did for review and ratings, we investigate how the effect of recruiting fake reviewers changes with the length of the campaign duration. In the right panel of Figure 5, we plot the focal products sales rank by Facebook campaign duration quartiles. The first thing that emerges from this plot is that there is a clear difference in rank between products with different Facebook promotion lengths: products promoted for shorter periods have a higher sales rank. Second, we observe that the negative effect on sales rank is more pronounced for products running shorter campaigns, but also (and as expected) this effect lasts for shorter periods as the length of the campaign decreases.

**Keywords search position** So far we have shown that recruiting fake reviews improves ratings, reviews, and sales. One reason for observing higher sales is that higher ratings signal higher quality to consumers, who then are more likely to buy the product. A second reason that could drive sales is that products recruiting fake reviews will be ranked higher in the
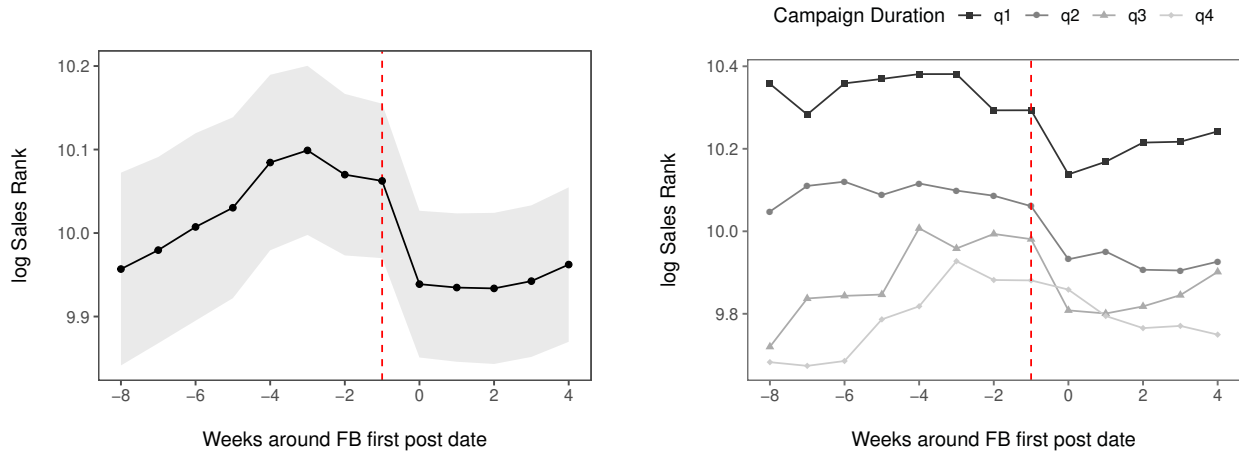
22

Figure 5: 7-day average sales rank before and after fake reviews recruiting begins overall products (left) and by campaign duration quartiles (right). The red dashed line indicates the first time we observe Facebook fake review recruiting.

Amazon search results due to them having higher ratings and more reviews (both factors that are likely to play a role in determining a product search rank). To investigate whether this is the case, in Figure 6 we plot the search position rank of products recruiting fake reviews. In the left panel, we plot the search rank for all products, while in the right panel we divide products based on their campaign duration. We observe a large drop in search position rank corresponding with the beginning of the Facebook promotions, indicating that products recruiting fake reviews improve their search position substantially. Moreover, this seems to be a lasting effect as the position remains virtually constant in the post-promotion period. Turning to the heterogeneous effect by campaign duration (right panel), we observe similar patterns across all groups with minimal differences in the effect of the Facebook campaign.

**Verified purchases and photos**   Next, we investigate the effect of recruiting fake reviewers on whether the review is written by someone who actually bought the product (Amazon 'Verified purchase" reviews) and the number of photos associated with the reviews. An important aspect of the market for fake reviews is that reviewers are compensated for creating realistic reviews, meaning they actually buy the product and can therefore be listed as a
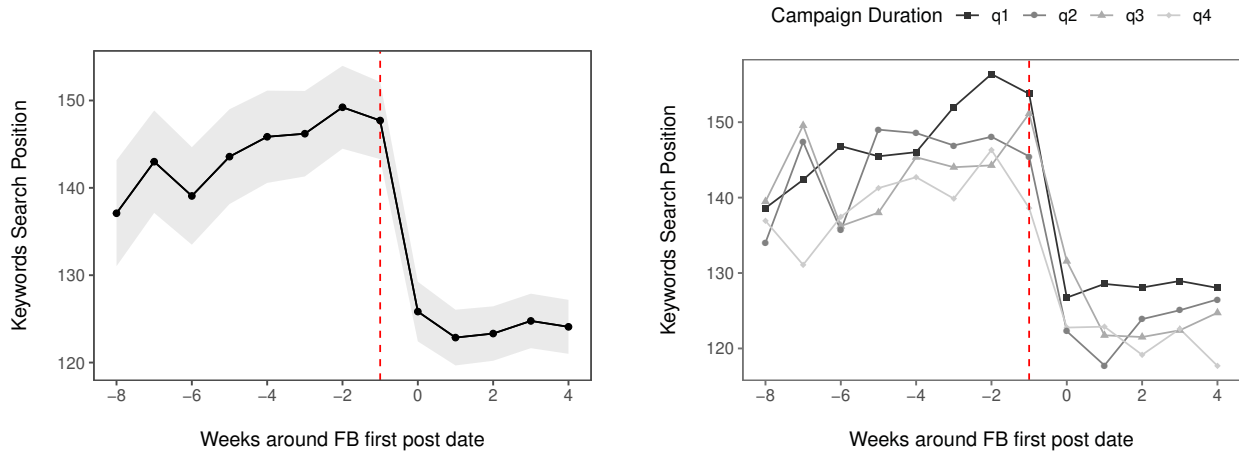
23

Figure 6: 7-day keywords search position before and after fake reviews recruiting begins overall products (left) and by campaign duration quartiles (right). The red dashed line indicates the first time we observe Facebook fake review recruiting.

verified reviewer, and they are encouraged to post long and detailed reviews. We plot these two quantities in Figure 7. In the left panel, we show changes in 7-day interval average verified purchase reviews. Despite being quite noisy in the pre-promotion period, the figure suggests that verified purchases increase with the beginning of the promotion. Turning to the number of photos (right panel) we observe a sharp increase that begins around interval -1 suggesting a positive effect associated with the beginning of the Facebook promotion.
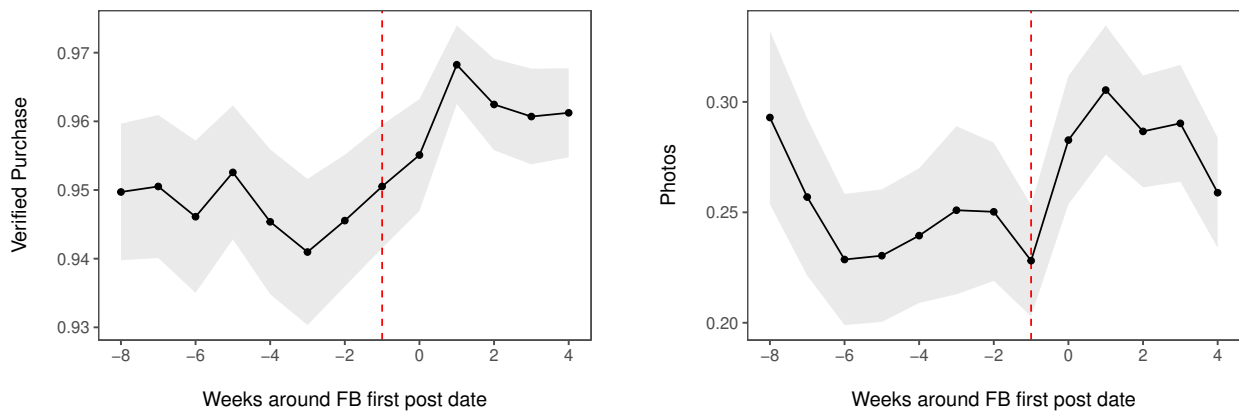


Figure 7: 7-day average verified purchase and number of photos before and after fake reviews recruiting begins. The red dashed line indicates the first time we observe Facebook fake review recruiting.

24

**Marketing activities** Finally, we investigate to what extent recruiting fake reviewers is associated with changes in other marketing activities such as promotions (sponsored listings and coupons). We plot these quantities in Figure 8. We observe a substantial negative change in prices (left panel) that persists for several weeks. We also observe a persistent increased use of sponsored listings suggesting that Amazon sellers complement the Facebook promotion with advertising activities. This result contrasts with Hollenbeck et al. (2019) which finds that online ratings and advertising are substitutes and not complements in the hotel industry, an offline setting with capacity constraints. Finally, we observe a small negative effect (albeit noisy) on the use of coupons.
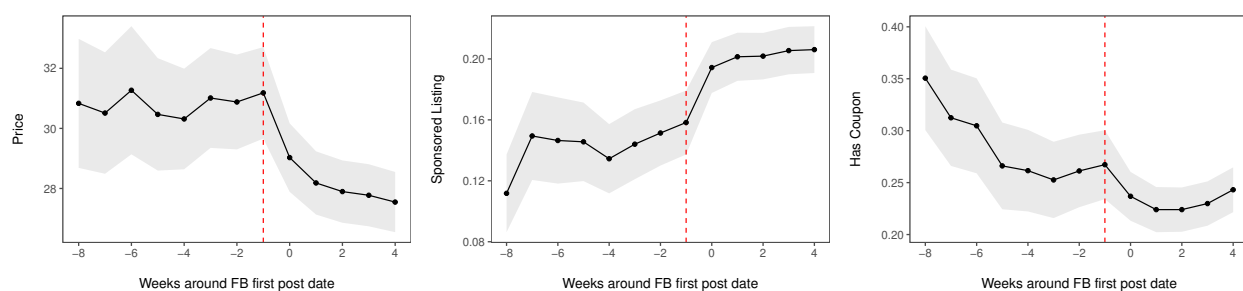


Figure 8: 7-day average sponsored listings and coupon. The red dashed line indicates the first time we observe Facebook fake review recruiting.

## 4.1 Regression Results

We have so far shown the effects of fake reviews visually. We now show the same results in a regression context in order to test whether the changes in outcomes we observe are statistically meaningful when a full set of fixed effects is included as well as to quantify the size of these changes.

We take the interval [-8,4] weeks and regress each outcome variable on a dummy for the time period from 0 to 2 weeks afterward, as well as an additional dummy for the time period after that. This divides up our sample into three periods: a before period, a period in which short-term effects should be present, and a period in which longer-term effects should be present. In each case we include year-week and product fixed effects.

25

The results for each variable are shown in Table 3. Consistent with our visual analysis, we see significant short-term improvements in average rating, number of reviews, sales, and search position (keyword rank). The increase in weekly average rating is roughly .07 stars. We also see significantly higher use of sponsored listings in this period and a significant increase in the share of reviews that are from verified purchases. There are also positive coefficients for the long-term dummy for the number of reviews and search position, confirming that these are long-lasting effects.

Table 3: Short-term Effects of Recruiting Fake Reviews

| Weeks After FB Post: | (1) Avg. Rating | (2) log Reviews | (3) log Sales Rank | (4) log Keyword Rank | (5) Sponsored | (6) Coupon | (7) log Photos | (8) Verified | (9) log Price |
|---|---|---|---|---|---|---|---|---|---|
| 0-2 | 0.069* | 0.238*** | −0.095*** | −0.316*** | 0.029* | −0.019 | 0.023** | 0.008* | −0.008* |
| | (0.027) | (0.020) | (0.027) | (0.034) | (0.011) | (0.015) | (0.009) | (0.004) | (0.004) |
| >2 | 0.039 | 0.111*** | −0.003 | −0.265*** | 0.031* | −0.054** | −0.000 | 0.008 | −0.011* |
| | (0.033) | (0.027) | (0.039) | (0.040) | (0.014) | (0.020) | (0.011) | (0.006) | (0.005) |
| N | 9563 | 9563 | 9516 | 6058 | 6131 | 6131 | 9563 | 9563 | 6093 |
| $R^2$ | 0.31 | 0.70 | 0.85 | 0.75 | 0.63 | 0.65 | 0.33 | 0.27 | 0.99 |

*Note:* All specifications include product and year-week FE. Cluster robust standard errors (at the product level) in parentheses.
*Significance levels:* * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Overall, we observe that when sellers purchase fake reviews there is an immediate and substantial increase in the number of reviews they receive and average ratings. Additionally, these products increase their use of marketing activities such as sponsored listings at this time, and the net effect of these are a large increase in their sales that persist for several weeks. In the next section, we document the long-term trends in these outcome variables after sellers stop buying fake reviews.

## 4.2   Causal Effect of Fake Reviews on Sales

The results presented so far are descriptive and cannot be interpreted as entirely causal effects. There are two concerns. The first is that sellers buying fake reviews may time these fake reviews around unobserved shocks to demand, either positive or negative. While the product fixed effects included in the results presented in Table 3 capture time-invariant
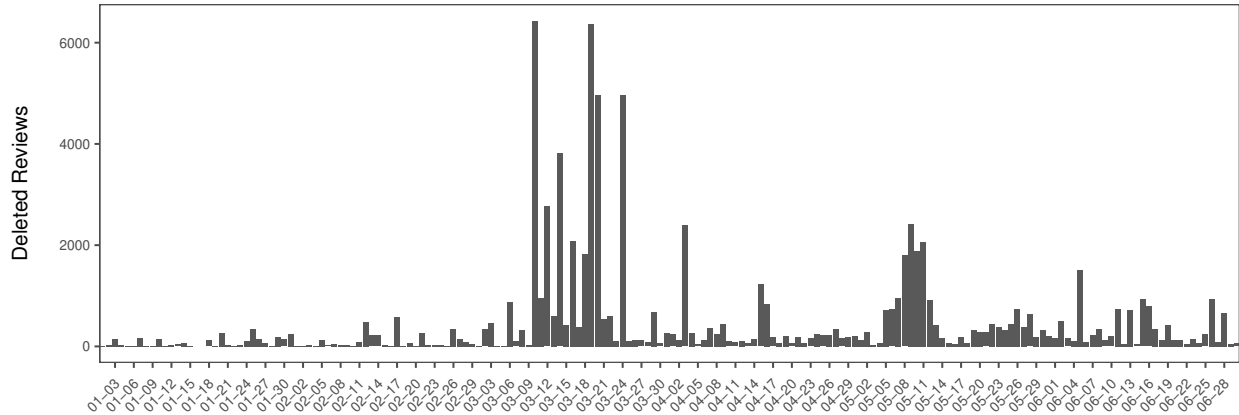
Figure 9: Amazon deleted reviews by date

unobserved heterogeneity they would not capture these shocks. The second is that we observe sellers cut prices and increase advertising at the same time they recruit fake reviews, making it difficult to isolate the effect of fake reviews on sales. While it may be clear that fake reviews do increase sales because we observe millions of products finding it worthwhile to recruit them, we are still be interested in isolating and measuring this effect to understand its magnitude.

To accomplish this we take advantage of an event that occurred during our sample period that provides a clean measurement of the effects of fake review recruiting. As we discuss at length in section 6, Amazon deletes a large number of reviews. Figure 9 shows the amount of review deletion over time during our sample period. There is one occasion in March 2020 when Amazon undertakes a large-scale purge of reviews with much higher rates of deletion than normal.[9] Assuming sellers had no foresight that these review purge was about to be undertaken, a subset of the sellers who recruited fake reviews had the misfortune of doing so during or just before the review purge occurred. By comparing their outcomes to those of products recruiting fake reviews at other times we can isolate the effect of fake review recruiting on sales.

We refer to the products who recruited fake reviews just before or during the review purge

---

[9]There is another spike in review deletion in May of 2020 but it affects substantially fewer reviews and is not as long-lasting.

27

as control products and all other products who recruited fake reviews as treated products. Our analysis requires three assumptions to hold to identify a causal effect. First, control products were otherwise similar to the treated products. Second, the control products took similar actions on prices and advertising as treated products. Third, the review purge was effective at preventing the control products from acquiring fake reviews. We analyze each of these assumptions and then show the estimated effect of fake review recruiting on sales.

We start by taking the midpoint date of the review purge, which is March 15th, and defining our set of control products as all products whose first observed Facebook post is in the interval [-2,1] weeks around this date. This gives us 46 control products who recruited before or during the peak of the review purge. We then use the window defined by [-8,4] weeks around this date but outside that window to define our set of comparison treated products. This provides 805 treated products.

We begin by testing the first assumption and show our results in Table 4. While the control set of products are younger, and hence have less cumulative reviews, the products are otherwise not significantly different and have quite similar cumulative average ratings and number of weekly reviews.

Table 4: Comparison of Treated and Control Products

|                     | Controls | Treated | Difference |
| ------------------- | -------- | ------- | ---------- |
| Age                 | 7.81     | 11.6    | $-3.81^*$  |
| Avg Ratings         | 4.32     | 4.13    | 0.19       |
| Cum. Avg. Ratings   | 4.37     | 4.35    | 0.021      |
| log Reviews         | 1.74     | 1.83    | $-0.099$   |
| log Cum. Reviews    | 4.10     | 4.66    | $-0.56^*$  |
| log Price           | 3.29     | 3.13    | 0.16       |
| Coupon              | 0.27     | 0.28    | $-0.0074$  |
| Verified            | 0.95     | 0.98    | $-0.032$   |
| log Photos          | 0.14     | 0.17    | $-0.027$   |

*Note:* Averages for non-cumulative variables are computed at the interval level for the period [-8,-2] intervals.

*Significance levels:* * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

28

We then estimate a standard difference-in-differences (DD) regression comparing outcomes for these products before and after they recruit fake reviews and comparing the treated products to the control products. The regression takes the following form:

$$y_{it} = \beta_1 \text{ Treated}_i + \beta_2 \text{ After}_{it} + \beta_3 \text{ Treated}_i \times \text{After}_{it} + \alpha_i + \tau_t + X'_{it}\gamma + \epsilon_{it}, \qquad (7)$$

where *Treated* is a dummy defined as we describe above and *After* is a dummy for the period after the first observed Facebook post. $\alpha$ are product fixed effects to account for time-invariant product characteristics that could be correlated with the outcome, and $\tau$ are week fixed effects to account for time-varying shocks to the outcome that affects all products (e.g., holidays). As it is common in DD analyses, we include treatment-specific weekly trends as an additional safeguard against non-parallel trends. The coefficient $\beta_2$ measures the effect of fake review recruiting for control products and the object of interest, $\beta_3$ measures the difference of this effect for treated products. We estimate the regression in equation 7 using OLS and clustering standard errors at the product level. We report the results for seven different outcomes in Table 5.

Table 5: DD Estimates

| | (1) log Cum. Reviews | (2) Sponsored | (3) Coupon | (4) log Price | (5) log Sales Rank | (6) log Keyword Rank |
|---|---|---|---|---|---|---|
| After | −0.001 | 0.011 | −0.005 | 0.000 | 0.166 | −0.193 |
| | (0.053) | (0.043) | (0.060) | (0.009) | (0.109) | (0.138) |
| After × Treated | 0.379*** | 0.019 | −0.018 | −0.007 | −0.272* | −0.129 |
| | (0.060) | (0.045) | (0.062) | (0.011) | (0.115) | (0.143) |
| N | 9563 | 6131 | 6131 | 5648 | 9516 | 6058 |
| R$^2$ | 0.92 | 0.63 | 0.65 | 0.99 | 0.85 | 0.75 |

*Note:* All specifications include product and year-week FE, and a treatment specific time-trend. Cluster-robust standard errors (at the product level) in parentheses.
*Significance levels:* * p<0.05, ** p<0.01, *** p<0.001.

Column 1 shows the results for the log of total number of reviews. The purpose is to assess whether the review purge did in fact prevent the control products from increasing their number of reviews through fake review recruiting. We see a small but insignificant increase in

reviews for control products and a large (38%) and significant increase for treated products. Next, columns 2 through 4 show results for the use of sponsored listings, coupons, and for log of price. This comparison allows us to assess the assumption that fake review recruiters behave the same regardless of whether their recruiting period coincided with the review purge. We see an insignificant increase in sponsored listings, a small decrease in the use of coupons, and virtually no changes in prices for the control group, and essentially no difference between those products and the treated products.

Finally, we assess the key outcome variables in column 5. We see that sales rank for the control products actually increases, although this increase is not statistically significant. The treated products sales rank decreases significantly, suggesting that the treatment of fake review recruiting is effective at increasing sales. Finally, Keyword rank (reported in column 6) shows also a large but insignificant benefit for the control products and a similarly sized additional benefit for the treated products, although the difference between them is not significant.

Overall these results suggest that the necessary assumptions hold and that increasing the number of reviews through fake review recruiting does have a significant causal effect on sales, holding other marketing actions and unobserved shocks fixed.

# 5   Long-term Outcomes from Buying Fake Reviews

In this section, we describe what happens after sellers stop buying fake reviews. Using the procedure described in section 2.1 we construct an end date for each product after they are no longer observed recruiting fake reviews. We are interested in using the long-term outcomes for two purposes. First, to assess whether buying fake reviews generates a self-sustaining increase in sales. Second, to evaluate the potential consumer harm from fake reviews.

If we observe these products continue to receive high organic ratings and high sales after recruiting fake reviews stops, we might conclude that fake reviews are a potentially helpful

way to solve the cold start problem of selling online with limited reputation. If, by contrast, we see declining sales and ratings and a large number of one-star reviews it suggests that the sellers buying fake reviews are using them to mask low quality products and deceive consumers. Receiving low ratings and a large share of one-star reviews would indicate that the actual quality of these products is lower than many customers expected at the time of their purchase. This would cause them harm, either because they paid a higher price than they would have if they had not been deceived by fake reviews, or because it caused them to buy a lower quality product than the closest alternative. This analysis is also important from the platform's perspective; an increase in one-star reviews would indicate fake reviews are a major problem since they reflect negative consumer experiences that should erode the sense of trust the platform's reputation system is meant to provide.

We therefore track the long-term trends for ratings, reviews, and sales rank. We also track the share of reviews that come with ratings of one star, the lowest possible rating, as an indicator of low product quality or consumers who feel they have been deceived into buying the product. Lastly, we perform a detailed text analysis of the post-recruiting one-star reviews to see if they are distinctive compared to other one-star reviews and, if so, what text features they are associated with.

Similar to Section 4, we partition the time around the last Facebook recruiting post date (day 0) in 7-day intervals, and plot the quantity of interest for four 7-day intervals before fake reviews recruiting stop (thus covering most of the period where products recruited fake reviews) and eight 7-day intervals after fake reviews recruiting starts. Doing so, we are comparing the Facebook promotion effect (negative intervals) with the post-promotion effect (positive intervals) where fake recruiting stopped.

**Ratings and Reviews**   The long term effects of fake reviews on ratings and reviews are shown in Figure 10. We observe that the effect of buying fake reviews is fairly short. After one-to-two weeks from the end of the Facebook promotion, both the number of reviews and
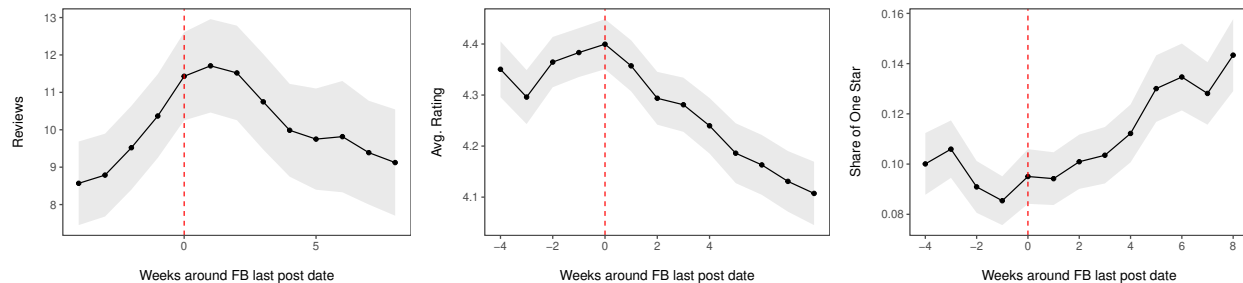
31

Figure 10: 7-day average number of reviews, average ratings, and average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last time we observe Facebook fake review recruiting.

average ratings (left and middle panel, respectively) start to decrease substantially. The right panel of Figure 10 clearly explains why the average rating is decreasing. The share of one-star reviews starts to increase considerably once recruiting fake reviews stops. Interestingly, these products end up having average ratings that are significantly worse than when they started recruiting fake reviews (approximately interval -4).

We next plot these outcomes for different duration of the Facebook campaign in Figure 11. Focusing on reviews (left panel), we observe that while all products experience a reduction in the number of reviews they receive, the decrease for products with shorter campaign duration is less pronounced. In fact, these products seem to continue to receive more reviews than when the Facebook promotion started. Instead, products with longer campaign duration are in a worse situation with the number of reviews going back to pre-Facebook campaign levels. We observe similar patterns for average ratings and share of one-star reviews (although all products seem to go back to levels worse than the pre-Facebook promotion period). These findings suggest that there might be a negative correlation between the quality of the product promoted and how long they are promoted.

Finally, we explore the long-term effect on the share of one-star reviews for different types of products. It may be the case that while one-star reviews increase after fake review purchases stop, certain products are able to retain high ratings. For example, new products (i.e., products with few reviews or that have been listed on Amazon for a brief period of time) might use fake reviews to bootstrap their reputation that then they can sustain if these
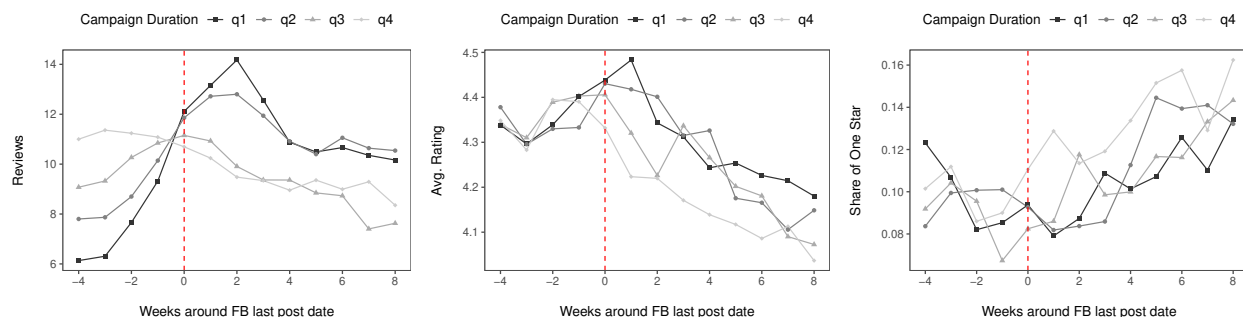
Figure 11: 7-day average number of reviews, average ratings, and average share of one-star reviews before and after fake reviews recruiting stops, by Facebook campaign duration quantities. The red dashed line indicates the last time we observe Facebook fake review recruiting.

products are high quality products.

To test this hypothesis, we replicate the right panel of Figure 10 but segmenting products by number of reviews and age. The left panel of Figure 12 shows how the share of one-star reviews changes for products with fewer than 50 reviews at the time they started recruiting fake reviews compared to all other products. The products with few reviews show a somewhat sharper increase in one-star ratings. The right panel of Figure 12 shows the same outcome, but for products that have been listed on Amazon for fewer than 60 days when they started recruiting fake reviews (very young products), compared to all other products. The young products have a much larger increase in one-star reviews compared to other products, with more than 20% of their ratings being one-star ratings two months after they stop recruiting fake reviews. Overall, these results do not support our hypothesis. Instead, they suggest that new products recruiting fake reviews are likely to be low quality products that use fake reviews to inflate their ratings and sales.

**Sales Rank**    Figure 13 shows the average log of sales rank. These two figures reveal several facts. First, sales decline substantially after the last observed Facebook post. This suggests that the effect of recruiting fake reviews is not long-lasting. It does not create a self-sustaining set of sales and positive reviews, in other words.
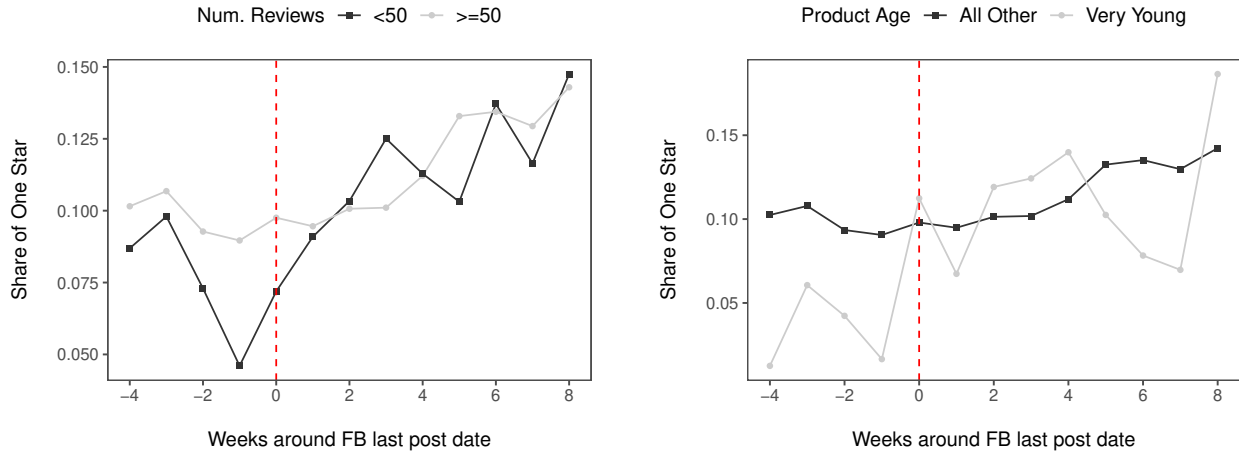
33

Figure 12: 7-day average sales rank before and after fake reviews recruiting stops for sub-samples of products with few reviews and new products (posted less than 60 days). The red dashed line indicates the last time we observe Facebook fake review recruiting.
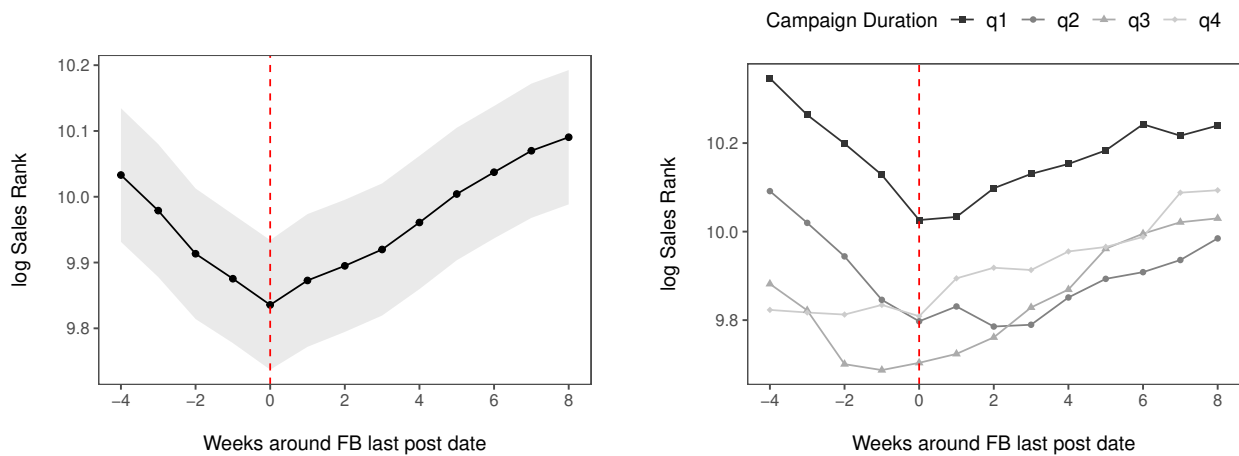


Figure 13: 7-day average sales rank before and after fake reviews recruiting stops overall focal products (left) and by campaign duration quartiles (right). The red dashed line indicates the last time we observe Facebook fake review recruiting.

**Keywords search position**    Figure 14 shows the average keywords search position over all products (left panel) and by campaign duration (right panel). We observe that after the Facebook campaign stops, the downward trend in search position stops but does not substantially reverse even after 2 months. Therefore all products enjoy better a more prominent ranking in keyword searches for a relatively long period after fake review recruiting stops. Turning to the figure by campaign duration (right panel), we observe a similar pattern that
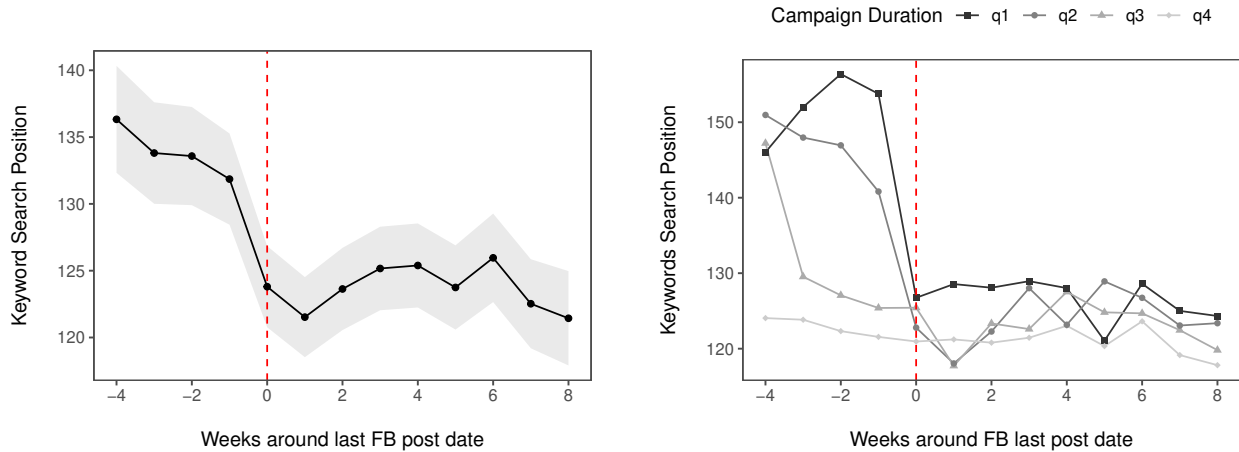
34

Figure 14: 7-day keywords search position before and after fake reviews recruiting stops overall products (left) and by campaign duration quartiles (right). The red dashed line indicates the last time we observe Facebook fake review recruiting.

confirms the almost steady state of search position rank after fake reviews recruiting stops.

The relatively stable and persistent increase in search position suggests that this measure may have a high degree of inertia. After an increase in sales and ratings cause a product's keyword rank to improve it does not decline quickly even when sales are decreasing. This also suggests that the decrease in sales shown in Figure 13 does not come about from reduced product visibility but from the lower ratings and increase in one-star reviews. Finally, while we demonstrate in the next section that Amazon deletes a large share of reviews from products that recruit fake reviews it does not punish these sellers using the algorithm that determines organic keyword rank. This could therefore serve as an additional policy lever for the platform to regulate fake reviews.

## 5.1   Text Analysis

So far, we have shown increases in the number and share of one-star reviews to provide evidence that consumers are harmed by the fake reviews that Amazon sellers buy through Facebook groups. Here, we provide additional evidence in support of this hypothesis by analyzing the text of these reviews.

35

We use state-of-the-art machine learning algorithms to analyze and compare the text of one-star reviews received by fake review products after the Facebook campaign ended with a random sample of one-star reviews of comparable products. The goal of this analysis is two-fold. First, we want to check whether negative reviews of the focal products are distinctive compared to those of competitors; second, if they are indeed different, we want to identify the text features that differentiate them. It may be the case, for instance, that one-star reviews increase after any sales spike and this is not a phenomenon specific to fake reviews. By analyzing the text we test if the negative reviews received by fake review products are distinctive from regular one-star reviews, and we use the text features to discern why they are distinctive and if they indicate harm to consumers.

We start by sampling 5,000 one-star reviews for each product type: those recruiting fake reviews and their competitors, defined as products showing up on the same results page for a keyword search. Then, we train a text-based classifier to predict whether each one-star review is from a product recruiting fake reviews or not. Following standard practice, we split the dataset into an 80% training sample and a 20% test sample. We present the results using a Naive Bayes classifier based on tf-idf. Depending on the configuration of the classifier (we can change the number of text features used by the classifier by removing very rare and very popular words), we achieve an accuracy rate that varies between 63% and 75% and a ROC-AUC score that varies between 69% and 83%.[10] These results suggest that the text of the reviews is sufficiently distinctive for the classifier to distinguish between the two types of one-star reviews. In other words, despite the products themselves being highly similar and the reviews having the same star-rating, the one-star reviews for products that had recruited fake reviews used a significantly different set of terms from one-star reviews for products that did not recruit fake reviews.

We next look at what are the most predictive text features for distinguishing the two product types. In Table 6, we report the top-30 features for the model achieving an accuracy

---

[10]Other types of classifiers led to similar performance.

rate and ROC-AUC score of 63% and 69%, respectively. What emerges from this table is that one-star reviews written for products recruiting fake reviews are predicted by text features mostly related to product quality (qualiti, stop work, work, etc.), value/price (waste money, money, disappoint, etc.), and past reviews (review) that may have deceived consumers into buying these products; instead, competitors' one-star reviews are predicted by text features mostly related to idiosyncratic product characteristic (second attach, fade, reseal, etc.). Overall, these results add further evidence that consumers who bought products that recruited fake reviews felt deceived in thinking that the products were of higher quality than they really were.

Table 6: Top-30 Most Predictive Features of the Text Classifier

| Products... | Text Features |
| --- | --- |
| recruiting fake reviews | work, product, money, return, use, time, stop, wast, month, would, like, wast money, charg, even, broke, stop work, week, disappoint, good, back, light, first, tri, bought, qualiti, review, turn, batteri, recommend, great |
| not recruiting fake reviews | reseal, command, bang, fixtur, apart piec, septemb, product dont, fade, ignit, use never, use standard, terrier, compani make, desktop, love idea, wifi connect, bead, solar panel, inexpens, within year, return sent, compani product, second attach, pure, cycl, thought great, solar charg, blame, bought march, price paid |

## 5.2 Which product characteristics predict consumer harm?

Next, we take the number and share of one-star reviews as indicators of low quality products and plausible consumer harm, and study what product characteristics predict increases in these variables after fake review recruiting stops. To do so, we start by measuring changes

37

in these variables between the post and pre-fake review recruiting periods using a 20-week window (10 weeks for each period). Additionally, we allow for possible lagged effects by excluding the 4-week window around the first Facebook post. The changes are thus defined using weeks [-12,-2] for the pre-fake review recruiting period and [2,12] for the post-fake review recruiting period.

We then regress these changes on product characteristics at the time of the first observed fake review post. We include product age, price, category, total number of reviews, sales rank, whether the seller is known to be located in China, and when the post occurred. This last variable is broken into seasons which we describe as Spring (March through May), Summer (June through September), Fall (October through December), and Winter (January through February).

These regressions are meant to capture correlations only since the decision to purchase fake reviews is endogenous and related to product characteristics. However, these correlations are informative and can show which characteristics predict larger or smaller consumer harm from fake reviews. We report these results in Table 7. We observe that the estimates are not sensitive to the definition of consumer harm (either the number or the share of 1-star reviews). We find that younger, lower ranked, and more expensive products see significantly larger increases in 1-star reviews after they stop recruiting fake reviews. There are also significantly larger increases for products that recruit fake reviews in the Spring and Fall. Products sold by Chinese sellers are not associated with significantly higher consumer harm.

# 6 Amazon Response

In this section we provide evidence on the extent to which Amazon is aware of the fake review problem and what steps it is taking to remove these reviews.

While we cannot observe reviews that are filtered by Amazon's fraud detection practices and never made public, by collecting review data on a daily and bi-monthly basis we can

Table 7: Long-Term Outcomes After Recruiting Fake Reviews

| | (1)<br>Change in # 1 Star Reviews | (2)<br>Change in Share of 1 Star Reviews |
|---|---|---|
| log Age | −0.243*** | −0.037*** |
| | (0.063) | (0.011) |
| log Cum. Reviews | −0.060 | 0.004 |
| | (0.039) | (0.007) |
| log Price | 0.171*** | 0.034*** |
| | (0.052) | (0.009) |
| log Sales Rank | −0.211*** | −0.004 |
| | (0.031) | (0.005) |
| China | 0.091 | −0.011 |
| | (0.070) | (0.012) |
| Spring | 0.356*** | 0.003 |
| | (0.086) | (0.015) |
| Summer | 0.353 | −0.030 |
| | (0.202) | (0.034) |
| Fall | 0.352** | 0.014 |
| | (0.119) | (0.020) |
| Arts, Crafts, & Sewing | −0.169 | −0.054 |
| | (0.409) | (0.070) |
| Baby | −0.538 | 0.015 |
| | (0.437) | (0.075) |
| Beauty & Personal Care | 0.320 | −0.051 |
| | (0.328) | (0.056) |
| Clothing, Shoes & Jewelry | 0.493 | 0.013 |
| | (0.596) | (0.102) |
| Electronics, Computers & Accessories | −0.036 | −0.046 |
| | (0.330) | (0.056) |
| Health & Household | 0.264 | −0.054 |
| | (0.331) | (0.057) |
| Home & Kitchen | 0.175 | −0.072 |
| | (0.326) | (0.056) |
| Industrial & Scientific | −0.019 | 0.005 |
| | (0.516) | (0.088) |
| Office Products | −0.087 | −0.109 |
| | (0.376) | (0.064) |
| Patio, Lawn, & Garden | 0.553 | −0.017 |
| | (0.356) | (0.061) |
| Sports & Outdoors | 0.085 | −0.011 |
| | (0.347) | (0.059) |
| Toys & Games | −0.267 | −0.162* |
| | (0.438) | (0.075) |
| Automotive | 0.104 | 0.012 |
| | (0.378) | (0.065) |
| Pet Supplies | 0.107 | −0.029 |
| | (0.345) | (0.059) |
| Tools & Home Improvement | 0.060 | −0.048 |
| | (0.337) | (0.057) |
| N | 869 | 869 |
| $R^2$ | 0.12 | 0.063 |

*Note:* Robust standard errors in parentheses.
*Significance levels:* * p<0.05, ** p<0.01, *** p<0.001.

39

observe if reviews are posted and then later deleted. We calculate the share of reviews that are deleted by comparing the full set of observed reviews from our daily scraper to the set of reviews that remain posted at the end of our data collection window. We find that for our set of products observed recruiting fake reviews the average share of reviews that are ultimately deleted is about 55%. This suggests that, to some extent, Amazon can identify fake reviews.

To further characterize Amazon's current policy we next analyze the characteristics of deleted reviews and the timing of review deletion.

**Characteristics of Deleted Reviews**   In Table 8, we report the mean and standard deviation for several review characteristics for existing reviews and deleted reviews, respectively. Following the literature on fake reviews, we focus on characteristics that are often found to be associated with fake reviews. Specifically, we focus on whether the reviewer purchased the product through Amazon (verified purchase), review rating, number of photos associated with the review, whether the reviewer is part of Amazon's "Early Reviewer Program", i.e. is one of the first users to write a review for a product the length of the review title, and the length of the review.[11]

We find that deleted reviews have higher average ratings than non-deleted reviews. This is driven by the fact that the vast majority of deleted reviews are five-star reviews (see Figure 15).

Deleted reviews are also associated with more photos, shorter review titles, and longer review text. In general, we might expect longer reviews, those that include photos, and those from verified purchases to be less suspicious. The fact that these are more likely to be deleted suggests that Amazon is fairly sophisticated in targeting potentially fake reviews.[12] Finally, we find no difference for whether the review is associated with a verified purchase

---

[11]For more details about the "Early Reviewer Program", we refer the reader to `https://smile.amazon.com/gp/help/customer/display.html?nodeId=202094910`

[12]This result contrasts with Luca and Zervas (2016), who find that longer reviews are less likely to be filtered as fake by Yelp.

or tagged as "Amazon Earlier Reviews".[13]

Table 8: Summary Statistics of Existing Reviews

|  | All Reviews Mean | Deleted Reviews Mean | t-stat |
|---|---|---|---|
| Verified purchase | 0.96 (0.19) | 0.97 (0.17) | 10.37 |
| Review rating | 4.41 (1.24) | 4.57 (1.10) | 34.05 |
| Number of photos | 0.22 (0.75) | 0.28 (0.85) | 27.65 |
| Early reviewer | 0.01 (0.08) | 0.00 (0.00) | $-18.01$ |
| Title length | 19.81 (15.15) | 16.94 (15.40) | $-54.01$ |
| Review length | 225.07 (242.11) | 240.15 (233.50) | 17.10 |

Standard deviations in parentheses.

**When Are Reviews Deleted?**   Finally, we analyze when Amazon deletes fake reviews for focal products. We do so by plotting the number of products for which reviews are deleted over time relative to the first Facebook post, i.e., the beginning of the buying of fake reviews. To do so, we partition the time in days around the Facebook first post, and then plot the number of products for which reviews are deleted. Because our focal products have different campaign durations, we do this analysis by campaign duration quartiles. Figure 16 shows the results of this analysis. What clearly emerges from this figure is that Amazon starts deleting reviews for more products after the Facebook campaign begins and often it does so only after the campaign terminated. Indeed, it seems that most of the review deletion happens during the period covering the two months after the first Facebook post date, but most campaigns are shorter than a month. A simple calculation suggests that reviews are deleted only after a quite large lag (when compared with the duration of the Facebook campaign). The mean

---

[13]We find that Amazon does not delete any reviews tagged as "Amazon Earlier Reviews" potentially because Amazon process to identify and select early reviewers drastically reduces the possibility of these reviews to be fake.
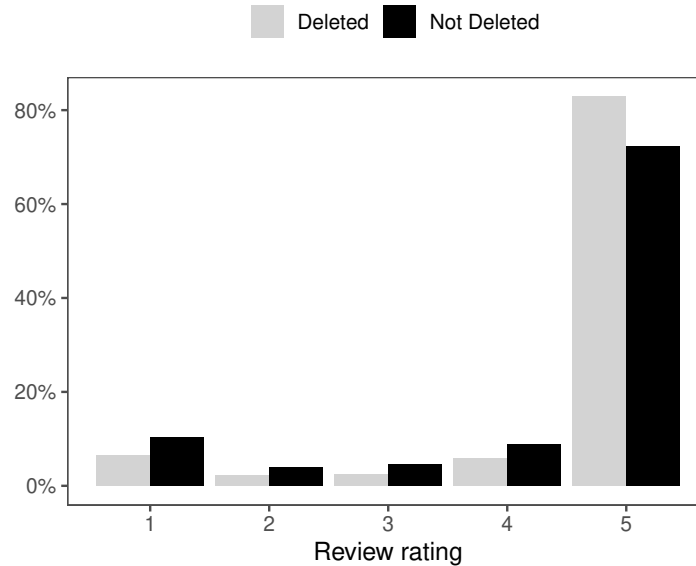
Figure 15: Histogram of Review Ratings of Reviews

time between when a review is posted and when it is deleted is over 100 days, with a median time of 53 days.

This analysis suggests the deleted reviews may be well-targeted at fake reviews, but that there is still a significant lag between when the reviews are posted and when they are deleted; and this lag allows sellers buying fake reviews to enjoy the short-term benefits of this strategy discussed in Section 4.

# 7    Discussion and Conclusions

Fake reviews have become commonplace for online sellers to manipulate their reputation on online platforms. In this paper, we study the market for fake reviews of one of the world's largest e-commerce platforms, Amazon. We study how fake reviews affect seller outcomes both in the short-term and in the long-term. These two analyses allow us to study both the immediate effectiveness of fake reviews and to understand whether these reviews are harming consumers and online platforms or not.

We find that the Facebook promotion is highly effective at improving several sellers'
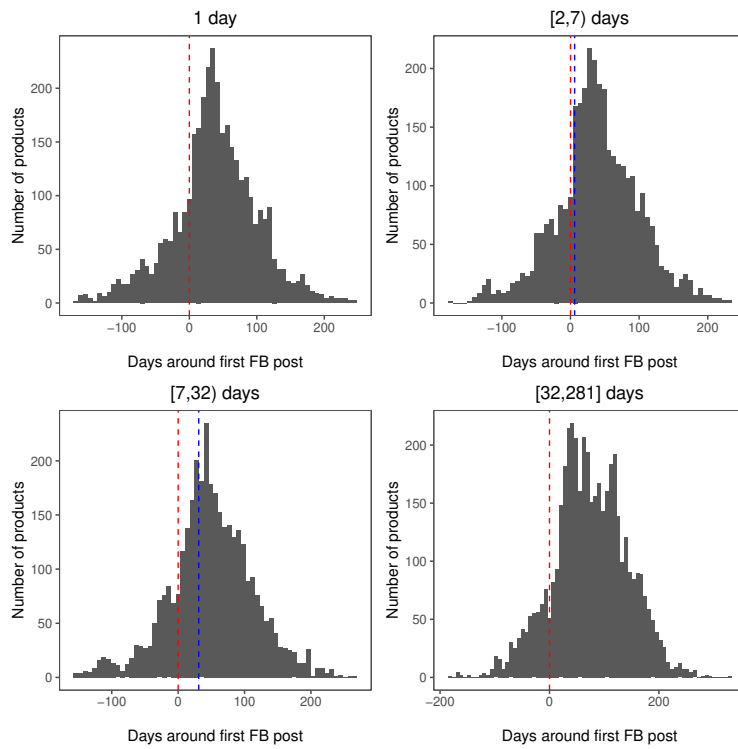
42

Figure 16: Number of products for which reviews are being deleted over time relative to the first Facebook post date. The red dashed line indicates the first time we observe Facebook fake review recruiting, and the blue dashed line indicates the last time we observe Facebook fake review recruiting.

43

outcomes such as number of reviews, ratings, search position rank, and sales rank, in the short-term. However, these effects are often short-lived as many of these outcomes return to pre-promotion levels a few weeks after the fake reviews recruiting stops. In the long run, this boost in sales does not lead to a positive self-sustaining relationship between organic ratings and sales and both sales and average ratings fall significantly once fake review recruiting ends. Fake reviews are not used efficiently by sellers to solve a "cold-start" problem, in other words.

This decrease in ratings once sellers stop buying fake reviews and the large increase in the share of one-star reviews suggest that consumers who bought these products were harmed. They either overpaid for the true quality of the product or bought a different, lower quality product than they would have in the absence of rating manipulation. In addition to harming consumers, rating manipulation likely harms honest sellers and the platform's reputation itself. If large numbers of low-quality sellers are using fake reviews, this also decreases the signal value of high ratings and causes consumers to be skeptical of new but highly rated products, which may ultimately make it harder for new, high-quality sellers to successfully enter the market and reduce innovation.

Firms are continuously improving and perfecting their manipulation strategies and this area is evolving at a very fast pace so that findings that were true only a few years ago, or strategies that could have worked in the past to eliminate fake reviews, might be outdated today. This is why studying and understanding how firms create fake reviews continue to be an extremely important topic of research for both academics and practitioners. As a testament to this Amazon claims to have spent over $500 million in 2019 alone and employed over 8,000 people to reduce fraud and abuse on its platform.[14]

Our last finding is therefore to document how Amazon responds to sellers recruiting fake reviews. We find that Amazon responds by deleting reviews at a very high rate. Moreover, we find that Amazon's response is quite sophisticated. The timing of review deletion suggests

---

[14]See: `https://themarkup.org/ask-the-markup/2020/07/21/how-to-spot-fake-amazon-product-reviews`

that it is able to identify which sellers and reviews are likely fake despite these reviews being very similar to real ones. However, while review deletion seems well targeted, there is a large lag between these reviews being posted and then deleted. In practice, this means that currently, Amazon is not able to eliminate the short-term profits from these reviews or the consumer harm we see expressed in the large number of one-star reviews.

We can also document that Amazon does not use other potential policy levers at its disposal to regulate fake reviews. We do not observe in our sample that Amazon either deletes many products or bans sellers as a result of them manipulating their ratings. We also do not observe punishment in the products' organic ranking in keyword searches. This keyword ranking stays elevated several months after fake review recruiting has ended, even as Amazon finds and deletes many of the fake reviews. Reducing product visibility in keyword rankings at the time fake reviews are deleted could potentially turn fake reviews from a profitable endeavor into a highly unprofitable one.

It is not obvious whether Amazon is simply under-regulating rating manipulation in a way that allows this market to continue to exist at such a large scale, or if it is assessing the short-term profits that come from the boost in ratings and sales and weighing these against the long-term harm to the platform's reputation. Quantifying these two forces is therefore an important area of future research.

Because consumers are harmed, even if Amazon is optimally regulating fake reviews according to its own objectives, the externality may be scope for greater action by public regulators like the FTC. This could take the form of trying to identify sellers who buy fake reviews and fining or punishing them in other ways. But because there are millions of these sellers and most of them are located in China, identifying and punishing them is likely to be impractical. Instead, regulators may be justified in pushing Amazon to use the policy levers at its disposal to prevent sellers from deceiving and harming consumers using fake reviews.

# References

Ananthakrishnan, U., Li, B., and Smith, M. (2020). A tangled web: Should online review portals display fraudulent reviews? *Information Systems Research*, forthcoming.

Anderson, E. and Simester, D. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51.

Cabral, L. and Hortacsu, A. (2010). The Dynamics Of Seller Reputation: Evidence From Ebay. *Journal of Industrial Economics*, 58(1):54–78.

Chiou, L. and Tucker, C. (2018). Fake news and advertising on social media: A study of the anti-vaccination movement.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science*, 52(10):1577–1593.

Einav, L., Farronato, C., and Levin, J. (2016). Peer-to-peer markets. *Annual Review of Economics*, 8(1):615–635.

Glazer, J., Herrera, H., and Perry, M. (2020). Fake reviews. *The Economic Journal*, forthcoming.

Gordon, B., Jerath, K., Katona, Z., Narayanan, S., Shin, J., and Wilbur, K. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, 85(1):7–25.

Hollenbeck, B. (2018). Online reputation mechanisms and the decreasing value of chain affiliation. *Journal of Marketing Research*, 55(5):636–654.

Hollenbeck, B., Moorthy, S., and Proserpio, D. (2019). Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science*, pages 793–811.

Li, X., Bresnahan, T. F., and Yin, P.-L. (2016). Paying incumbents and customers to enter an industry: Buying downloads.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

Mayzlin, D., Y., D., and Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *The American Economic Review*, 104:2421–2455.

Milgrom, P. and Roberts, J. (1986). Prices and Advertising Signals of Product Quality. *Journal of Political Economy*, 94:297–310.

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy*, 78(2):311–329.

Proserpio, D. and Zervas, G. (2016). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. *Marketing Science*.

Rao, A. (2018). Deceptive claims using fake news marketing: The impact on consumers.

Rao, A. and Wang, E. (2017). Demand for "healthy" products: False claims and ftc regulation. *Journal of Marketing Research*, 54.

Rhodes, A. and Wilson, C. M. (2018). False advertising. *The RAND Journal of Economics*, 49(2):348–369.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8(1):321–340.

Wu, Y. and Geylani, T. (2020). Regulating deceptive advertising: False claims and skeptical consumers. *Marketing Science*, 39(4):669–848.

Yasui, Y. (2020). Controlling fake reviews.