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‡

December, 2021 §

*sherry.he.phd@anderson.ucla.edu
†brett.hollenbeck@anderson.ucla.edu
‡davide.proserpio@marshall.usc.edu

§Authors listed in alphabetical order. Corresponding author: Brett Hollenbeck. We thank the Morrison Center for Marketing Analytics for generous funding. We also thank Elizabeth Paz, Jason Lei, Yoon Sang Moon, and Rachel Ziffer for excellent research assistance and seminar participants at the Harvard Business School, University of Arizona, Duke University Fuqua School of Business, Frankfurt School of Finance & Management, Kellogg School of Management at Northwestern, Olin Business School at Washington University in St Louis, Singapore Management University, Stanford Graduate School of Business, D'Amore-McKim School of Business at Northeastern, USC Marshall marketing workshop, UCLA Anderson marketing workshop, Virtual Quant Marketing Seminar, 2020 Conference on Digital Experimentation at MIT, 2020 WISE Conference, and the Toulouse Conference on Digital Economics for helpful suggestions.

Abstract

We study the market for fake product reviews on Amazon.com. Reviews are purchased in large private groups on Facebook and other sites. We hand collect data on these markets and then collect a panel of data on these products' ratings and reviews on Amazon, as well as their sales rank, advertising, and pricing policies. We find that a wide array of products purchase fake reviews, including products with many reviews and high average ratings. Buying fake reviews on Facebook is associated with a significant but short-term increase in average rating and number of reviews. We exploit a sharp but temporary policy shift by Amazon to show that rating manipulation has a large causal effect on sales. Finally, we examine whether rating manipulation harms consumers or whether it is mainly used by high-quality in a manner like advertising or by new products trying to solve the cold-start problem. We find that after firms stop buying fake reviews, their average ratings fall and the share of one-star reviews increases significantly, particularly for young products, indicating rating manipulation is mostly used by low-quality products.

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.[1] And yet online platforms have become some of the world's largest companies in part by effectively limiting these practices and earning consumer trust. The economics of platforms suggest a difficult trade-off between opening the platform to outside actors such as third-party sellers and retaining strict control over actions taken on the platform. Preventing fraudulent or manipulative actions is key to this trade-off.

One such practice is manipulating reputation systems with fake product reviews. Conventional wisdom holds that fake reviews are particularly harmful because they inject noise and deception into systems designed to alleviate asymmetric information, cause consumers to purchase products that may be of low quality, and 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).

We study the economics of rating manipulation and its effect on seller outcomes, consumer welfare, and platform value. Despite being illegal, we document the existence of large and active online markets for fake reviews.[2] Sellers post in private online groups to promote their products and pay customers to purchase them and leave positive reviews. 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

[1]Recent work has documented many examples including firms increasing their visibility in search rankings via fake downloads (Li et al., 2016), increasing revenue via bot-driven advertising impressions (Wilbur and Zhu, 2009; Gordon et al., 2021), manipulating social network influence with fake followers, manipulating auction outcomes, or defrauding consumers with false advertising claims (Rao and Wang, 2017; Chiou and Tucker, 2018; Rao, 2021).

[2]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, toward 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`.

research assistants into these groups to document which products are buying fake reviews and when.[3] We then track these products' outcomes on Amazon.com, including their reviews, ratings, prices, and sales rank. This is the first data of this kind, providing direct evidence on the fake reviews themselves and on the outcomes from buying fake reviews.

The mere existence of such a large and public market for fake reviews on the largest e-commerce platform presents a puzzle. Given the potential reputation costs, why does Amazon allow this? In the short run, platforms may benefit from allowing fake positive reviews if these reviews increase revenue by generating sales or allowing for higher prices. It is also possible that fraudulent reviews are not misleading on average if high-quality firms are more likely to purchase them than low-quality firms. They could be an efficient way for sellers to solve the "cold-start" problem and establish a good reputation. Indeed, Dellarocas (2006) shows that this is a potential 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 that demand increases convexly with respect to user rating. Given how ratings influence search results, it is plausible that this condition holds. Other attempts to model fake reviews have also concluded that they may benefit consumers and markets.[4] The mechanism is different, but intuitively 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 positive word-of-mouth once consumers have discovered their true quality. If fake reviews generate sales which, in turn, generate future organic ratings, a similar dynamic could play out. In this case, fake reviews may be seen as harmless substitutes for

[3] While technically the seller buys the fake reviews, not the product, because our analysis is done at the product level and sellers often have many products, for clarity we refer to products buying fake reviews.

[4] These attempts include (Glazer et al., 2020) and Yasui (2020). In addition, 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.

advertising rather than as malicious. Therefore, we are left with an empirical question as to whether or not to view rating manipulation as representing a significant threat to consumer welfare and platform reputations.

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? How effective are they at increasing sales? Does rating manipulation ultimately harm consumers 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 answered using the unique nature of our data.

We construct a sample of approximately $1,500$ products observed soliciting fake reviews over a nine-month period. We find a wide assortment of product types in many categories. Many products have a large number of reviews and few are new to Amazon. These products do not have especially low ratings, with an average rating slightly higher than comparable products we do not observe soliciting fake reviews. 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 than for brand name firms (Hollenbeck, 2018).

We then track the outcomes of these products before and after the buying of fake reviews. In the weeks after they start to purchase fake reviews, the number of reviews posted per week roughly doubles. The average rating and share of five-star reviews also increase substantially, as do search position and sales rank. The increase in average ratings is short-lived, with ratings falling back to the previous level within two to four weeks, but the increase in the weekly number of reviews, sales rank, and position in search listings remains substantially higher more than four weeks later. We also track outcomes after the last observed post soliciting fake reviews and find that the increase in sales is not self-sustaining. Sales begin to fall significantly right after the fake review campaign ends. New products with few reviews, which might be using fake reviews efficiently to solve the cold-start problem, see a larger increase in sales initially but a similar drop-off afterward.

We also document how the platform regulates fake reviews. We see that Amazon ultimately deletes a very large share of reviews. For the products in our data observed buying fake reviews, roughly half of their reviews are eventually deleted, but the deletions occur with an average lag of over 100 days, thus allowing sellers to benefit from the short-term boost in ratings, reviews, and sales.

Next, to understand how effective and profitable this practice is, we leverage review deletions to measure the causal effect of fake reviews on sales. Our previous results are descriptive, and the increase in sales we document could be attributed in part to factors other than fake reviews, include unobserved demand shocks, advertising, or price cuts. To isolate the effect of rating manipulation on sales, we take advantage of a short period in which Amazon mass deletes a large number of reviews. Products that purchased fake reviews just before this period do not receive the boost in positive reviews that other products buying fake reviews do, but they behave similarly otherwise, allowing us to use these products as a control group. Comparing outcomes across products, we find that rating manipulation causes a significant increase in sales.

Lastly, we turn to the question of whether rating manipulation is efficient or it harms consumers. To do so, we study reviews and ratings posted after the fake review purchases end. If the products continue to receive high ratings from consumers, it suggests that fake reviews are more like advertising and are mainly bought by high-quality products, potentially solving the cold-start problem. If, by contrast, ratings fall and they receive many one-star ratings, it suggests that consumers felt they were deceived into buying products whose true quality was lower than they expected at the time of purchase and, therefore, they overpaid or missed out on a higher quality alternative. While there is an inherent limitation in using ratings to infer welfare, we nevertheless find that the evidence primarily supports the consumer harm view. The share of reviews that are one-star increases by 70% after fake review purchases, relative to before. This pattern is especially true for new products and those with few reviews. Text analysis shows that these one-star reviews are distinctive and place a greater focus on product quality, further confirming that consumers were deceived.

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) use Yelp's review filtering algorithm as a proxy for fake reviews and find that these reviews are more common on pages for firms with low ratings, independent restaurants, and restaurants with more close competitors. Using lab experiments, Ananthakrishnan et al. (2020) show that a policy of flagging fake reviews but leaving them posted can increase consumer trust in a platform.

We contribute to this literature by documenting the actual market where fake reviews are purchased and the sellers participating in this market. This data gives us a direct look at rating manipulation, rather than merely inferring its existence. Our data on firm outcomes before and after rating manipulation allow us to understand the short- and long-term effectiveness of rating manipulation 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). A difficult tension has always existed in the literature on online reviews, coming from the fact that the reviews and ratings being studied may be manipulated by sellers. 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.

2 Data and Settings

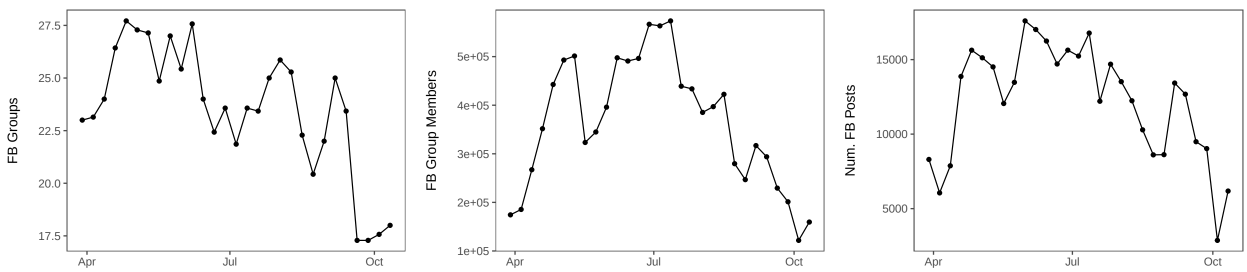
In this section, we document the existence and nature of online markets for fake reviews and discuss in detail the data collection process and the data we obtained to study rating manipulation and its effect on seller outcomes, consumer welfare, and platform value. We collected data mainly from two different sources, Facebook and Amazon. From Facebook, we obtained data about sellers and products buying fake reviews, while from Amazon we collected product information such as reviews, ratings, and sales rank 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, sellers create private Facebook 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 by searching for "Amazon Review." We begin by documenting the nature of these groups and then describe how we collect product information from them.

Discovering groups We collected detailed data on the extent of Facebook group activity from March 28, 2020 to Oct 11, 2020. Each day, we collected the Facebook group statistics for the top 30 groups by search rank. 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. Figure 1 shows the weekly average number of active groups, number of members,

Figure 1: Weekly average number of FB groups, members, and seller posts



and number of posts between April and October of 2020.[5]

Within these Facebook groups, sellers can obtain a five-star review that looks organic. Figure 2 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. The reviewer and seller then communicate via Facebook private messages. 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. The fact that the reviewer buys the product means that the Amazon review is listed as a "Verified Purchase" review and

[5]The total number of members and posts likely overstates the true amount of activity due to double-counting the same sellers and reviewers across groups.

reviewers are encouraged to leave lengthy, detailed reviews that include photos and videos to mimic authentic and organic reviews.[6] Finally, sellers recruit only reviewers located in the United States, with an Amazon.com account, and with a history of past reviews.

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. [7]

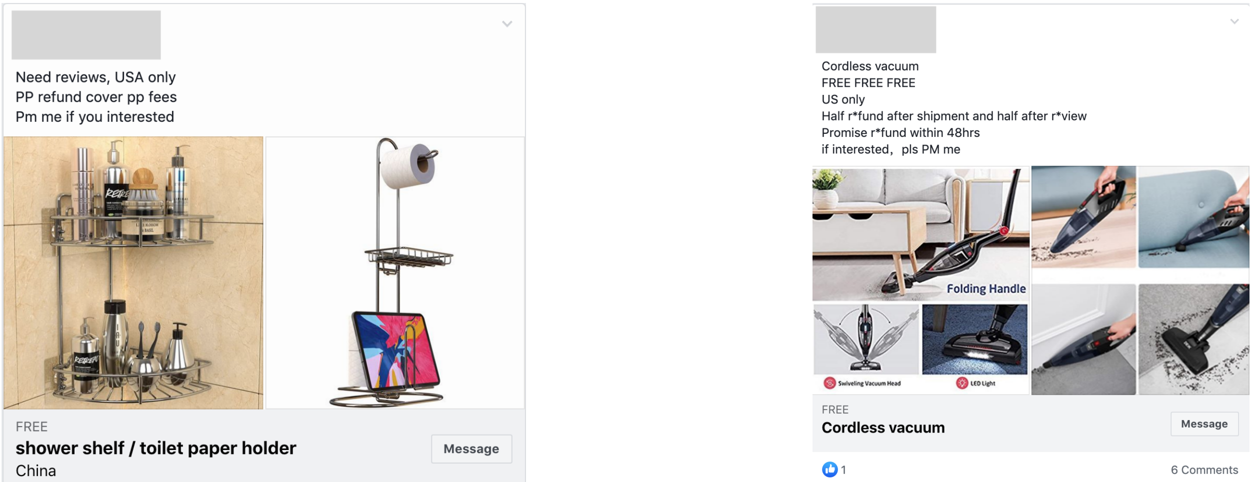
Discovering products We use a group of research assistants to discover products that are promoted. Facebook displays the posts in a group in an order determined by some algorithm that factors in when the post was made as well as engagement with the post via likes and comments. Likes and comments for these posts are relatively rare and so the order is primarily chronological. We directed our research assistants to randomize which products were selected by scrolling through the groups and selecting products in a quasi-random way while explicitly ignoring the product type/category, amount of engagement with the post, or the text accompanying the product photo.

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 2, the search words would be "shower self," "toilet paper holder," and "cordless

[6] 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 these.

[7] Amazon has at times allowed incentivized reviews and even has formally sponsored them through its Vine program and 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.

Figure 2: Examples of Fake Review Recruiting Posts



vacuum." After a research assistant successfully identifies the product, we ask them to document the 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 searching for the seller's Facebook name and 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 random 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

After identifying products whose ratings are manipulated, we collect data for these products on Amazon.com.

Search Results Data For each product buying fake reviews, we repeatedly collect all information from the keyword search page results, i.e., the list of products returned as a result of a keyword search query. This set of products is useful to form a competitor set for each focal product. We collect this information daily, including price, coupon, displayed rating, number of reviews, search page number, whether the product buys sponsored listings, and the product position in each page.

Review Data We collect the reviews and ratings for each of the products on a daily basis. For each review, we observe rating, product ID, review text, presence of photos, and helpful votes.

Additionally, twice per month we collect the full set of reviews for each product. The reason for this is that it allows us to measure to what extent Amazon responds by deleting reviews that it deems as potentially fake.

In addition to collecting this data for the focal products, we collect daily and twice-monthly review data for a set of 2,714 competitor products to serve as a comparison set. To do so, for each focal product we select the two competitor products who show up most frequently on the same search page as the focal product in the seven days before and seven days after their first FB post. The rationale is that we want to create a comparison set of products that are in the same subcategory as the focal products and have a similar search rank. We collect these products' reviews data from Aug 14th, 2020 to Jan 22rd, 2021.

Sales Rank Data We rely on Keepa.com and its API to collect sales rank data twice a week for all products. 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 an ordinal ranking of products.

2.3 Descriptive Statistics

Here, 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 sellers promote with fake reviews. On the one hand, we might expect these products to be primarily products that are new to Amazon.com with few or no reviews whose sellers are trying to jump-start sales by establishing a good 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.

Table 1 shows a breakdown of the top 15 categories and subcategories in our sample. Fake reviews are 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, and the most represented product in our sample, Humidifiers, only accounts for roughly 1% of products. Nearly all products are sold by third-party sellers.

Table 1: Focal Product Top 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

We observe substantial variation in the length of the recruiting period, with some products being promoted for a single day and others for over a month. The average length of the Facebook promotion period is 23 days and the median is six days.

In Table 2, 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. We observe that the focal products are significantly younger than competitor products, with a median age of roughly five months compared with 15 months for products not observed buying fake reviews. But with a mean age of 229 days, the products collecting fake reviews are not generally new to Amazon and without any reputation. Indeed, out of the 1,500 products we observe, only 94 solicit fake reviews in their first month.

Focal products charge slightly lower average prices than their competitors, having a mean price of $33 (compared with $45 for the comparison 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.

Turning to ratings, we observe that products purchasing fake reviews have, at the time of their first Facebook 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. Only 14% of focal products have ratings below four stars, compared with 19.5% for competitor products. Thus, it appears that products purchasing fake reviews do not seem to do so because they have a bad reputation. Although, we note that ratings may of course be influenced by previous unobserved Facebook campaigns.

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 with the set of competitor products, which has a median of 59 reviews and a mean of 451 reviews. Despite these differences, it seems that only a small share of the focal products have very few or no reviews. We also observe that the focal products have slightly lower sales

Table 2: Characteristics of Focal Products and Comparison Products

	Count	Mean	SD	25%	50%	75%
Displayed Rating						
Fake Review Products	1,315.0	4.4	0.5	4.1	4.5	4.8
All Products	203,480.0	4.2	0.6	4.0	4.3	4.6
Number of Reviews						
Fake Review Products	1,425.0	183.1	493.5	10.0	45.0	167.0
All Products	203,485.0	451.4	2,619.0	13.0	59.0	250.0
Price						
Fake Review Products	1,425.0	33.4	45.0	16.0	24.0	35.0
All Products	236,542.0	44.7	154.8	13.0	21.0	40.0
Sponsored						
Fake Review Products	1,425.0	0.1	0.3	0.0	0.0	0.0
All Products	236,542.0	0.1	0.3	0.0	0.0	0.0
Keyword Position						
Fake Review Products	1,425.0	21.4	16.1	8.0	16.0	33.0
All Products	236,542.0	28.2	17.3	13.0	23.0	43.0
Age (days)						
Fake Review Products	1,305.0	229.8	251.1	77.0	156.0	291.0
All Products	153,625.0	757.8	797.1	257.0	466.0	994.0
Sales Rank						
Fake Review Products	1,300.0	73,292.3	151,236.4	7,893.3	26,200.5	74,801.5
All Products	5,647.0	89,926.1	323,028.9	5,495.0	21,610.0	72,563.5

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 still be buying fake reviews via other (more private) channels, or they may avoid buying fake reviews altogether to avoid damages 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' names from Amazon and check the U.S. Trademark Office for records on each seller. We find a match for roughly 70% of

Table 3: Seller Characteristics

	Count	Mean	SD	25%	50%	75%
Focal Sellers						
Number of Products	660.0	23.9	83.9	3.4	7.8	15.2
Number of Reviews	642.0	176.9	297.0	34.0	81.2	201.1
Price	655.0	37.2	71.1	16.4	23.5	37.2
Seller Country						
Mainland China	798.0	0.8				
United States	112.0	0.1				
Hong Kong	13.0	0.0				
Japan	7.0	0.0				
Canada	6.0	0.0				

Note: This table shows information on seller characteristics, where the number of products, number of reviews and price variables are calculated as averages taken over all seller products. Variable counts differ based on the structure of Amazon seller pages making data collection impossible for some sellers. The number of observations for seller country is calculated at the product level.

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. The distribution of sellers by country-of-origin and other seller characteristics are shown in Table 3. This table shows that most sellers sell fewer than 15 products, with a median 7.8 products. Their products tend to have fewer than 200 reviews, similar to the focal products. The sellers' other products are also priced similarly to the focal products.

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 fake review marketplace works, and briefly show the costs and benefits of buying fake reviews. We start by focusing on the costs the sellers incur when buying a fake review.

First, to buy one 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, τ 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 one 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 one fake review:

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

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

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

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

$$Benefit = Q_o P(\lambda - c) \tag{4}$$

where again c refers to Amazon's commission from the sale. Setting equations 3 and 4 equal allows us to calculate the break-even number of organic sales Q_o^{BE}. This is the number of extra incremental sales necessary to exactly justify buying one 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:

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

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:

- $\tau = .0656$[8]
- $F_{PP} = 2.9\%$
- Amazon commission c varies by category but is either 8% or 15% in almost all cases.[9]

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

$$Q_o^{BE} = \frac{1.175 - \lambda}{\lambda - .08} \tag{6}$$

Thus the break-even level of incremental sales needed to justify buying one fake review is a simple expression of a product's price-cost margin. It is clear that products with larger markups require fewer incremental organic sales 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.

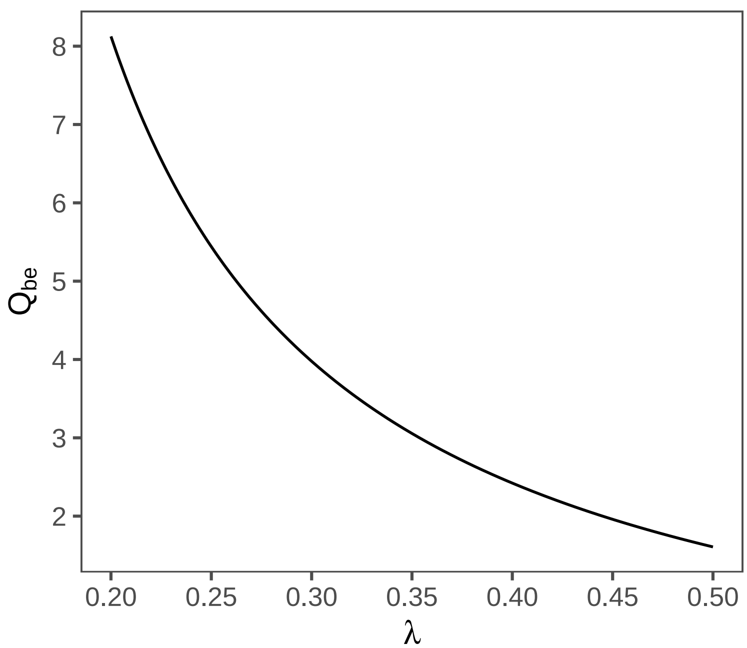
Figure 3 plots equation 6 where the X-axis is λ and the Y-axis is Q_o^{BE}. It shows that, for products with relatively low markups, the break-even number of organic sales approaches 10, but for products with relatively high markups, this number is below 1.

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

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

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 one additional organic sale to profit from a fake review purchase.

Figure 3: Organic sales needed to justify one fake review



Second, this is especially the case for lower quality products with larger markups. For 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 one 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 Descriptive Results on Product Outcomes After Buying Fake Reviews

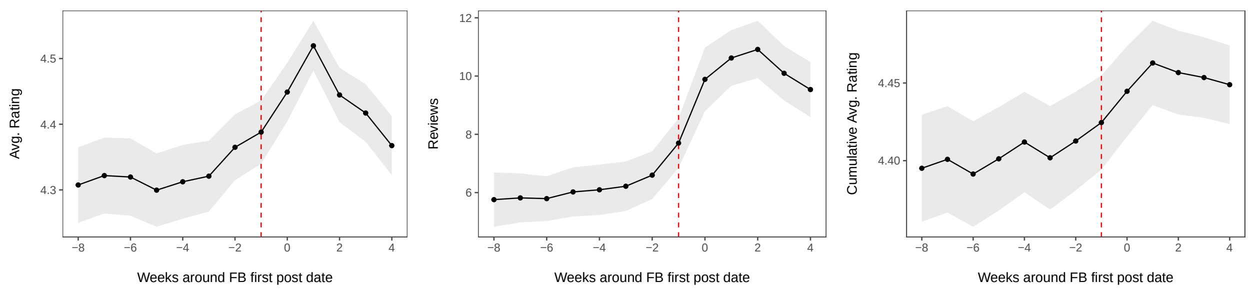
In this section, we quantify the extent to which buying fake reviews is associated with changes in average ratings, number of reviews, and sales rank, as well as other marketing activities such as advertising and promotions. To do so we take advantage of a unique feature of our data in that it contains a detailed panel on firm outcomes observed both before and after sellers buy fake reviews. We stress that, in this section, the 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. We present results on the causal effects of fake reviews on sales outcomes in Section 5.

We first present results in the short term, i.e., immediately after sellers begin buying fake reviews for their listings. We then show results for the persistence of these effects after the recruitment period has ended. Finally, we show descriptive results on the extent to which Amazon responds to this practice by deleting reviews.

4.1 Short-term Outcomes After Buying Fake Reviews

We begin by quantifying the extent to which buying fake reviews is associated with changes in average ratings, reviews, and sales rank in the short term. To evaluate these outcomes, we partition the time around the earliest Facebook recruiting post date (day 0) in 7-day intervals.[10] We then plot outcomes for eight 7-day intervals before and four 7-day intervals after the first fake review recruitment post.

Figure 4: 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 last week of data before we observe Facebook fake review recruiting.



Ratings and reviews We first examine ratings and reviews. In the left panel of Figure 4 we plot the weekly average rating after rating manipulation begins. We see that, first, the average ratings increase by about 5%, from 4.3 stars to 4.5 stars at its peak. Second, this increase in rating is short-lived, and it starts dissipating just two weeks after the beginning of the fake review recruiting; 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 increases slightly 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 beginning date of the fake review recruitment fairly well because the largest change in outcome is visible after or on interval zero

Next, we turn to the number of reviews. In the middle panel of Figure 4, we plot the weekly average number of posted reviews. We observe that the number of reviews increases substantially around interval zero, nearly doubling, providing suggestive evidence that recruiting fake reviewers is effective at generating new product reviews at a fast pace.

[10]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).

Moreover, and differently from the average rating plot, the increase in the weekly number of reviews persists for more than a month. This increase in the number of reviews likely reflects both the fake reviews themselves and additional organic reviews that follow naturally from the increase in sales we document below. Finally, Figure 4 confirms that we are not able to capture the exact date at which the Facebook promotion started.

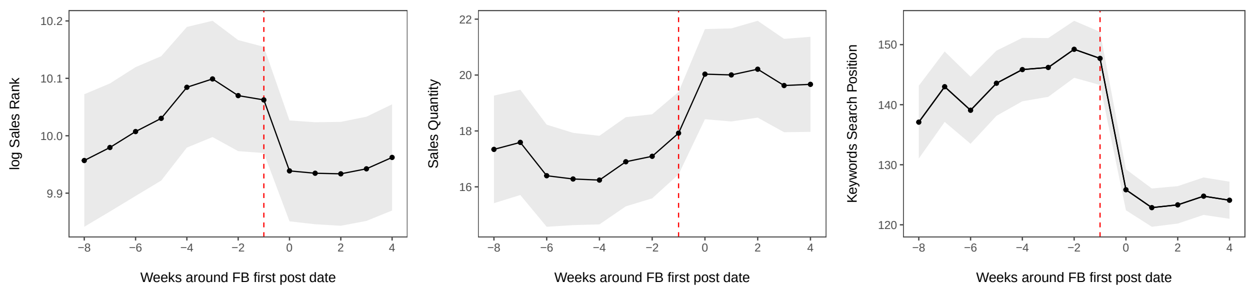
Does the increase in reviews lead to higher displayed ratings? To answer this question, in the right panel of Figure 4, we plot the cumulative average rating before and after the Facebook promotion starts. We observe that ratings increase and then stabilize for about two weeks, after which the increase starts to dissipate.

Sales rank In the left panel of Figure 5, we plot the average log of sales rank. The figure shows that the sales rank of these products increases between the intervals -8 and -3, meaning that rating manipulation typically follows a period when sales are falling. When the recruiting period begins, we observe a large increase in weekly sales (i.e. sales rank falls.) This increase is likely reflecting both the initial product purchases by the reviewers paid to leave fake reviews as well as the subsequent increase in organic sales that follow. The increase in sales lasts for at least several weeks.

The center panel of Figure 5 plots sales in units sold. Amazon does not display this metric but it is possible to measure sales in units for a subset of products and then estimate the relationship between rank and units. Appendix A describes how we collected this data and modeled the relationship, and more details are available in He and Hollenbeck (2020). We plot the observed sales and point estimates of estimated sales around the time of the first Facebook post and see a sharp increase in average units sold, from around 16 units per week to roughly 20.

Keyword search position So far we have shown that recruiting fake reviews is associated with improvements in ratings, reviews, and sales. One reason for observing higher sales may be that higher ratings signal higher quality to consumers, who then are more likely to buy

Figure 5: 7-day average sales rank before and after fake reviews recruiting begins (left), sales in units (center), and keyword search position (right) before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.

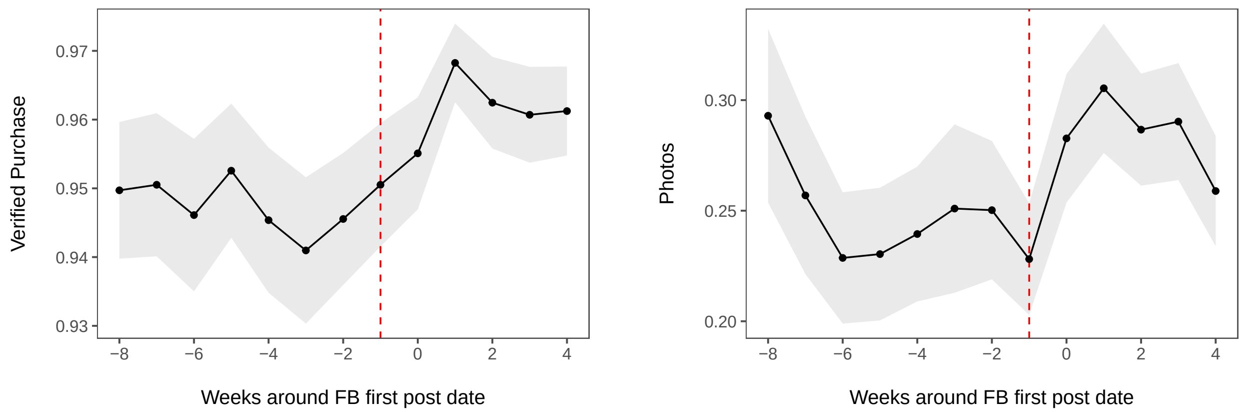


the product. A second reason is that products recruiting fake reviews will be ranked higher in the Amazon search results due to them having higher ratings and more reviews. To investigate whether this is the case, in the right panel of Figure 5 we plot the search position rank of products recruiting fake reviews. 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 change seems to be long-lasting as the position remains virtually constant for several weeks.

Verified purchases and photos An important aspect of the market for fake reviews is that reviewers actually buy the product and can therefore be listed as a verified reviewers. In addition, they are are compensated for creating realistic reviews, i.e, they are encouraged to post long and detailed reviews including photos and videos. In the left panel of Figure 6, we show changes in the average share of 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. In the right panel, we observe a sharp increase in the share of reviews containing photos.

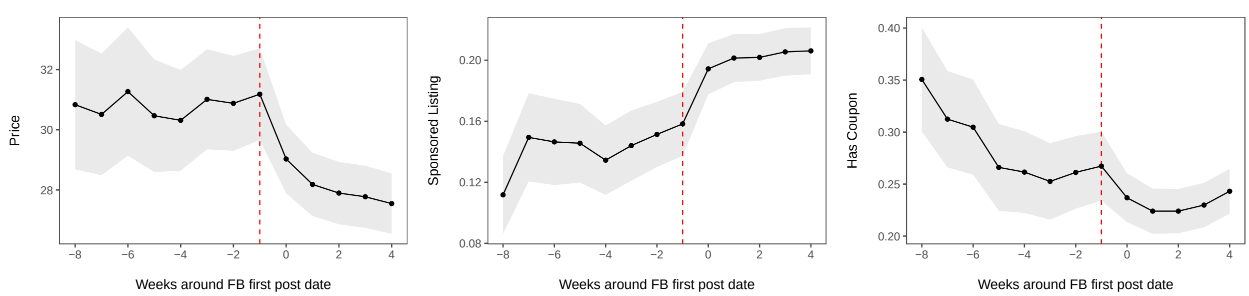
Marketing activities Finally, we investigate to what extent rating manipulation is associated with changes in other marketing activities such as promotions (rebates, sponsored listings, and coupons). We plot these quantities in Figure 7. We observe a substantial drop

Figure 6: 7-day average verified purchase and number of photos before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.



in prices (left panel) that persists for several weeks and an increase in the use of sponsored listings, suggesting that Amazon sellers complement the Facebook promotion with advertising activities. This result is in contrast with Hollenbeck et al. (2019) who find 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 (albeit noisy) change in the use of coupons.

Figure 7: 7-day average sponsored listings and coupon. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.



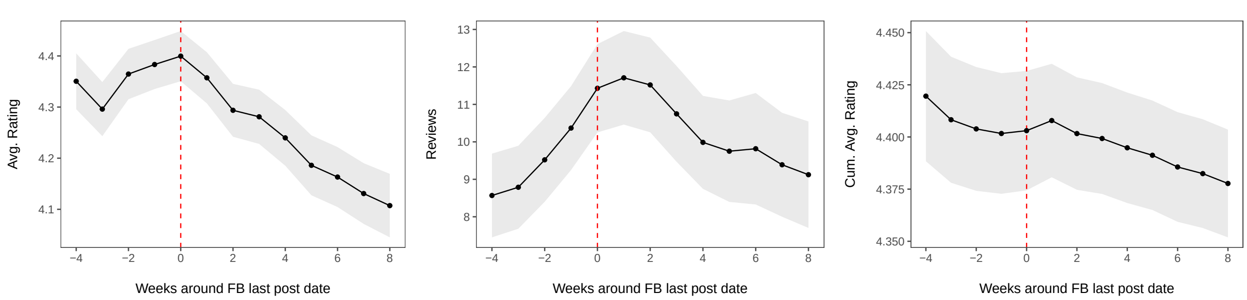
4.2 Long-term Outcomes After Buying Fake Reviews

In this subsection, we describe what happens after sellers stop buying fake reviews. We are particularly interested in using the long-term outcomes to assess whether rating manipulation generates a self-sustaining increase in sales or organic reviews. If we observe that these products continue to receive high organic ratings and have high sales after they stop recruiting fake reviews, we might conclude that fake reviews are a potentially helpful way to solve the cold-start problem of selling online with limited reputation.

We therefore track the long-term trends for ratings, reviews, and sales rank. Similar to Section 4.1, we partition the time around the last Facebook recruiting post date in 7-day intervals, and plot the outcomes for four weeks before fake reviews recruiting stop (thus covering most of the period where products recruited fake reviews) and eight weeks after fake reviews recruiting starts. Doing so, we compare the Facebook promotion period (negative intervals) with the post-promotion period (positive intervals).

Ratings and Reviews Long-term trends in ratings and reviews reviews are shown in Figure 8. We observe that the increase that occurs when sellers buy fake reviews is fairly short. After one to two weeks from the end of the Facebook promotion, both the weekly average rating and the number of reviews (left and middle panel, respectively) start to decrease substantially. The cumulative average rating (right panel) drops as well. Interestingly, these products end up having average ratings that are significantly worse than when they began recruiting fake reviews (approximately interval -4).

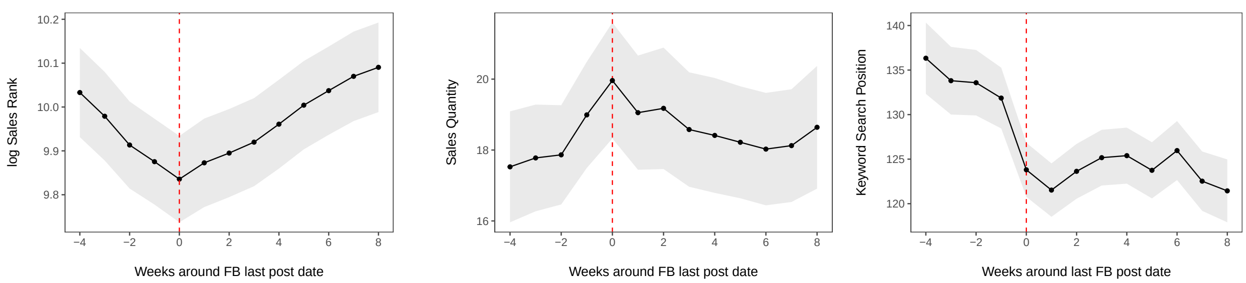
Figure 8: 7-day average number of average ratings, reviews, and average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last week of data in which we observe Facebook fake review recruiting.



Sales Rank The left panel of Figure 9 shows the long-term trend in the average log sales rank. It shows that sales decline substantially after the last observed Facebook post. This suggests that the increase associated with recruiting fake reviews is not long lasting as it does not lead to a self-sustaining set of sales and positive reviews.

The middle panel of Figure 9 shows sales in units, estimated using the procedure described in Appendix A. The result is consistent with sales rank, showing that sales peak during the week of the last Facebook post and subsequently decline.

Figure 9: 7-day average sales rank, sales in units, and keyword rank before and after fake review recruiting stops. The red dashed line indicates the last week of data in which we observe Facebook fake review recruiting.



Keyword search position The right panel of Figure 9 shows the long-term trend in average keyword search position. We observe that after the Facebook campaign stops, the downward trend in search position stops but does not substantially reverse even after two months. Therefore, products enjoy a better ranking in keyword searches for a relatively long period after fake review 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 causes 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 9 does not come from a 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, the inertia in keyword rank suggests that Amazon 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.

4.3 Regression and Heterogeneity Analysis

We have so far shown the outcomes associated with recruiting fake reviews visually. Appendix B shows the same results in a regression context 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 for all products and specific subgroups of products.

Consistent with our visual analysis, we see significant increases in average rating, number of reviews, sales, and search position (keyword rank) after fake review recruiting begins. 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. The regression results also confirm that the changes in the number of reviews and search position are especially persistent. Regression results also confirm the visual analysis that shows that average ratings, number of reviews, sales and keyword position all fall after fake review recruiting ends.

Using the regression framework we are also able to test if outcomes differ upon relevant dimensions of product heterogeneity. We are particularly interested in understanding whether there are larger changes in ratings, reviews, and sales for new products with few reviews, as these may buy fake reviews to alleviate the cold-start reputation problem. Regression results shown in Table 10 of Appendix B show that, in the short-term period after the first Facebook post for fake reviews, these new products do see their sales increase by a much larger margin than for regular products. They also get a larger increase in number of reviews but do not see an increase in weekly average rating.

After they stop buying fake reviews, we find that these products' ratings fall even further than for regular products, but that their increase in number of weekly reviews is more persistent. The persistence of their increase in weekly reviews corresponds to a larger and more persistent increase in sales. These results combine to suggest that rating manipulation is associated with especially positive outcomes for this type of product.

In a separate web appendix, we also analyze how outcomes vary along other dimensions of product heterogeneity, namely: high vs low-priced products, search vs experience goods, durable vs non-durable products, and products by category.

4.4 Amazon's Response

In this subsection, 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 twice-monthly basis, we can 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 with the set of reviews that remain posted at the end of our data collection window. We find that for the set of products observed recruiting fake reviews, the average share of posted reviews that are ultimately deleted is about 43%, compared to 23% for products not observed recruiting fake reviews. 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 4, we report the mean and standard deviation for several review characteristics for deleted and non-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 10). 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 reviews 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 or tagged as "Amazon Earlier Reviews."[13]

Table 4: Comparing deleted and non-deleted reviews characteristics

	Deleted Reviews	Non-deleted Reviews
Verified purchase	0.98	0.96
	(0.16)	(0.20)
Review rating	4.65	4.24
	(0.98)	(1.37)
Number of photos	0.35	0.19
	(0.93)	(0.72)
Early reviewer	0.00	0.01
	(0.00)	(0.11)
Title length	9.81	21.08
	(13.94)	(13.80)
Review length	236.73	198.75
	(222.88)	(231.68)

Note: Standard deviations in parentheses.

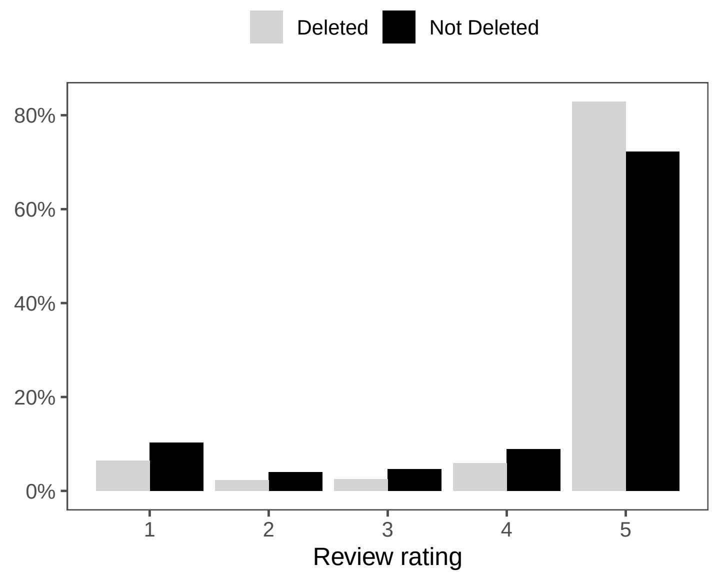
When Are Reviews Deleted? Finally, we analyze when Amazon deletes fake reviews for focal products. We do so by plotting the number of products for which reviews are deleted

[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.

[13]We find that Amazon does not delete any reviews tagged as "Amazon Earlier Reviews" potentially because Amazon's process to identify and select early reviewers drastically reduces the possibility of these reviews being fake.

Figure 10: Rating distribution for deleted and non deleted reviews

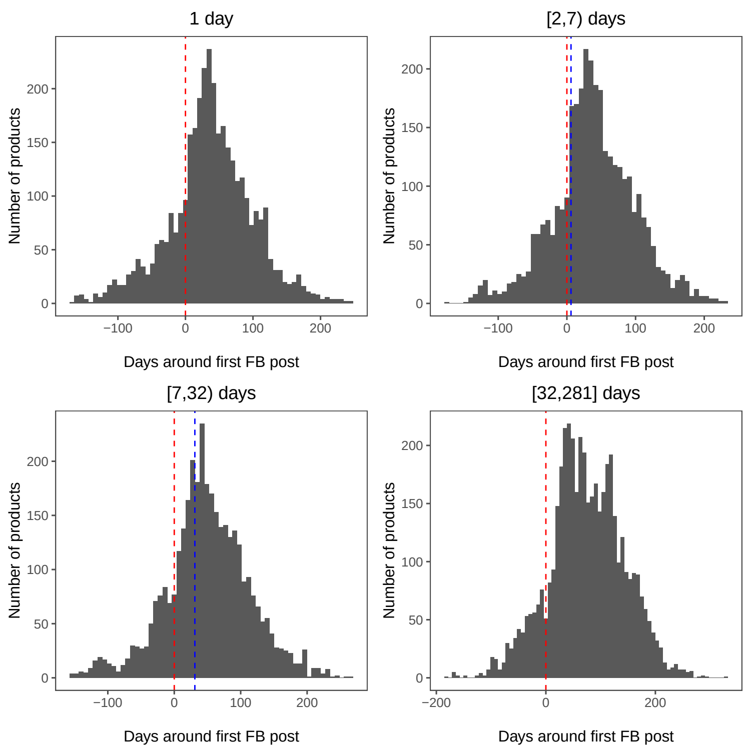


over time relative to the first Facebook post, i.e., the beginning of the buying of fake reviews. To do so, we partition the time in days around the first Facebook post and then plot the number of products for which reviews are deleted. Because products recruit fake reviews for different time periods, we perform this analysis by segmenting products based on the quartiles of campaign duration. Figure 11 shows the results of this analysis.

What emerges from this figure is that Amazon starts deleting reviews for more products after the Facebook campaign begins (red-dashed line) and often it does so only after the campaign terminated (blue-dashed line). 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. The mean 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 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 4.1. In the next section, we show that there is one time period in our data during which Amazon's deletion policy changes significantly; we use this period to identify the causal effects of fake reviews on sales.

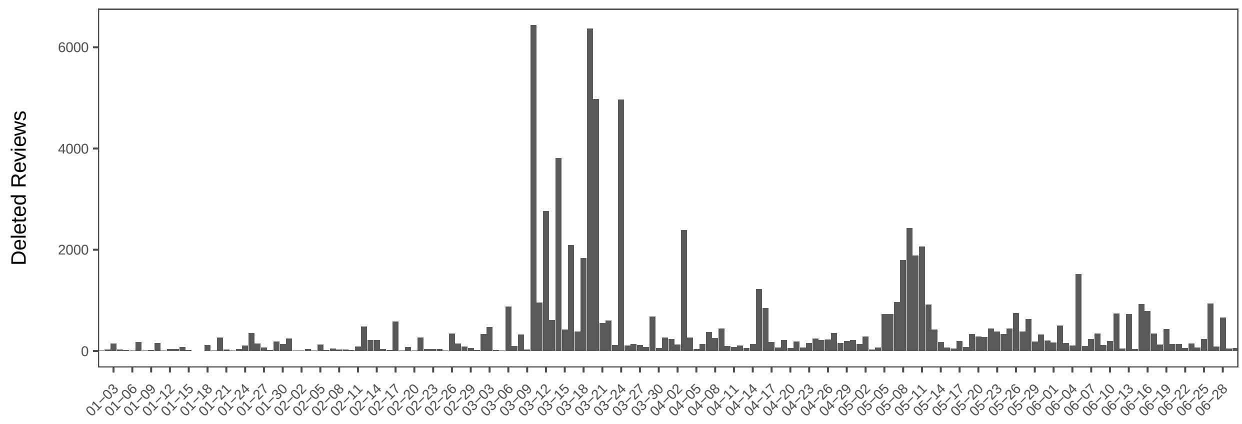
Figure 11: Number of products for which reviews are being deleted over time relative to the first Facebook post date. The red dashed line indicates the first time we observe Facebook fake review recruiting, and the blue dashed line indicates the last time we observe Facebook fake review recruiting.



5 The Causal Effect of Fake Reviews on Sales

In this section we measure the size of the effect of fake reviews on sales. The results in the previous section are descriptive and may not provide a valid estimate of the effect size. There are two concerns. The first is that sellers buying fake reviews may time these purchases around unobserved shocks to demand, either positive or negative. While product fixed effects capture time-invariant unobserved heterogeneity, they would not capture these shocks. The second concern is that many sellers change prices and advertising at the same time they recruit fake reviews, making it difficult to isolate the effect of fake reviews on sales. To

Figure 12: Amazon deleted reviews by date



overcome this, we exploit a temporary change in Amazon policy that allows us to isolate and measure the causal effect of fake review recruiting on sales. This measurement is useful to understand the effects that fake reviews have on sales and to establish that this is a profitable strategy for sellers.

To accomplish this, we take advantage of an event that occurred during our sample period. As we discussed in Section 4.4, Amazon deletes a large number of reviews, albeit after a lag. Figure 12 shows the amount of review deletion over time for the products seen buying fake reviews. There is one occasion during mid-March 2020 when Amazon undertakes a large-scale purge of reviews with much higher rates of deletion than normal and without a lag.[14] Assuming sellers had no foresight that this 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. Therefore, the products of these unlucky sellers should have no (or a much smaller) increase in positive reviews after they recruited fake reviews compared to the other products. We thus refer to these as control products and all other products that recruited fake reviews at different times as treated products. We can therefore employ a difference-in-differences (DD) strategy that compares the change in sales of treated and control products to estimate the size of the effect of rating manipulation.

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

In our case, the DD identification strategy requires four assumptions to hold to identify a causal effect. First, Amazon should not have strategically selected the products for which reviews were deleted, i.e., control products should be similar to treated products in both observable and unobservable characteristics. Second, the review purge should be effective at preventing the control products from acquiring fake reviews. Third, treated and control products should not differ in their use of marketing activities that can affect sales. Fourth, the parallel trends assumption should hold, i.e., pre-treatment sales trends for treated and controls products should be similar. We start by presenting the empirical strategy setup, we then test each of the assumptions discussed above, and then provide estimates and robustness checks.

5.1 Empirical strategy setup

We start by taking the midpoint date of the review purge, which is March 15, 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 results in 78 control products. The 1,412 products whose sellers started recruiting fake reviews outside of this window is the set of treated products.

We then estimate a standard DD regression which takes the following form:

$$y_{it} = \beta_1 \text{ Treated}_i + \beta_2 \text{ After}_{it} + \beta_3 \text{ Treated}_i \times \text{After}_{it} + \alpha_i + \tau_t + X_{it}^{'}\gamma + \epsilon_{it}, \tag{7}$$

where y_{it} is the outcome of interest for product i at year-week t, $Treated_i$ is an indicator for whether product i is treated and $After_{it}$ is an indicator for the period after the first observed Facebook post for product i. α are product fixed effects to account for time-invariant product characteristics, and τ are year-week fixed effects to account for time-varying shocks to the outcome that affect all products (e.g., holidays). The coefficient β_2 measures the effect of fake review recruiting for control products, and the coefficient of interest, β_3, is the classical DD estimate which measures the difference in sales for treated products. We estimate the regression in Equation 7 using OLS and clustering standard errors at the product level.

5.2 Identification checks

Treated and control products are similar To test this assumption, we show that (1) treated and control products are similar in most of their observable characteristics, and (2) Amazon does not seem to select specific products with the review purge. In Table 5 we compare treated and control products over a large set of variables by taking the average over the period [-8,-2) weeks before the products begin to recruit fake reviews.[15] We find that they are largely similiar but that control products are older, with lower average weekly ratings, and more cumulative reviews.

Table 5: Comparison of Treated and Control Products

	Control	Treated	t-stat
Age	9.84	7.15	2.36*
Weekly Avg. Ratings	4.10	4.32	-2.07*
Cum. Avg. Ratings	4.32	4.43	-1.36
Weekly Reviews	5.21	5.78	-0.33
Cumulative Reviews	234.80	109.90	3.11**
Price	27.10	33.60	-1.38
Coupon	0.23	0.26	-0.37
Verified	0.92	0.93	-0.60
Number of Photos	0.25	0.26	-0.15
Category	41.60	40.20	0.42

Note: t-test for equality of means for treated and control units. Means are computed at the interval level for the period [-8,-2) weeks.
Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

To reduce concerns about differences between treated and control products that could affect the DD estimates, we employ Propensity Score Matching (PSM) (Rosenbaum and Rubin, 1983) to match treated and control products on the observable variables that are dif-

[15]We exclude weeks [-2,-1] because the analysis in Section 4.1 suggests that for some products, outcomes start to change up to two weeks before the first Facebook post.

ferent across treatment conditions, i.e., age, weekly average ratings, and cumulative reviews. To do so, for every product, we average these variables over the period [-8,-2) weeks and then implement PSM using the Gaussian kernel matching procedure with a bandwidth of 0.005, and imposing a common support, i.e., we drop treatment observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls.[16] We start with 1,412 treated and 78 control products and, after matching, we are left with 987 treated and 48 control products. We verify that PSM eliminates the imbalance between treated and control units by computing a weighted (using the PSM weights) t-test for equality of means of treated and control products. We report the results of this test in Table 6.[17]

Table 6: Comparison of Treated and Control Products after matching

	Control	Treated	t-stat
Age	7.78	7.63	0.10
Weekly Avg. Ratings	4.07	4.15	-0.48
Cum. Avg. Ratings	4.34	4.33	0.07
Weekly Reviews	5.11	7.24	-0.72
Cumulative Reviews	109.48	124.69	-0.39
Price	25.74	32.63	-1.20
Coupon	0.24	0.27	-0.30
Verified	0.94	0.95	-0.31
Number of Photos	0.22	0.24	-0.19
Category	43.30	39.32	0.75

Note: Weighted t-test for equalty of means for treated and control units. Means are computed at the interval level for the period [-8,-2) weeks.

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

Turning to Amazon's criteria of selecting which products' reviews are deleted, in Appendix C, we show that review deletion during the purge period is highly concentrated on individual reviewers and is not targeted at specific products.

[16]We choose a bandwidth that allowed for a good matching, meaning that there are no longer any statistically significant differences between treated and control units for the variables used for matching.

[17]In Appendix D, we show that our results are not sensitive to the type of matching algorithm used.

Manipulation Check Here we present evidence showing that the review purge creates a valid set of control products. To do so, the purge must prevent these products, who were observed attempting to buy fake reviews, from receiving the treatment of an increase in reviews. We do so by estimating Equation 7 with the outcome set to be the log of cumulative reviews. We report these results in column 1 of Table 7. As expected, $After$ is small and close to zero, suggesting that there is no increase in reviews for control products. However, the interaction coefficient $After \times Treated$, is positive and significant and suggests that the number of cumulative reviews for treated products increased by approximately 10% more than control products.

Table 7: Diff-in-Diff Estimates

	(1) log Cum. Reviews	(2) Sponsored	(3) Coupon	(4) log Price	(5) log Sales Rank
After	0.047	0.014	0.011	-0.003	0.198*
	(0.036)	(0.026)	(0.047)	(0.009)	(0.097)
After × Treated	0.099*	0.027	-0.031	0.006	-0.375**
	(0.048)	(0.032)	(0.046)	(0.013)	(0.116)
PSM Sample	Yes	Yes	Yes	Yes	Yes
N	12620	7477	7477	7417	11553
R^2	0.96	0.65	0.65	0.99	0.87

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.

Marketing activities are similar To investigate whether treated and control products' marketing activities are similar, we estimate Equation 7 for three different outcomes: (1) whether product i buys sponsored listings: (2) whether product i offers discounts through coupons; and (3) product i price. We report these estimates in columns 2-4 of Table 7. We do not observe any statistically significant change in sponsored listings, coupons, and price after the first Facebook post for both treated and control products. Therefore, the assumption about marketing activities being similar across treatment and control products is satisfied.

Parallel trends Finally, we test the parallel assumption. To do so we estimate the following Equation:

$$y_{it} = \beta_1 \text{ Treated}_i + \beta_2 \text{ After}_{it} + \lambda_k \text{ Treated}_i \times \text{Week}_{kit} + \alpha_i + \tau_t + X_{it}^{'}\gamma + \epsilon_{it}, \tag{8}$$

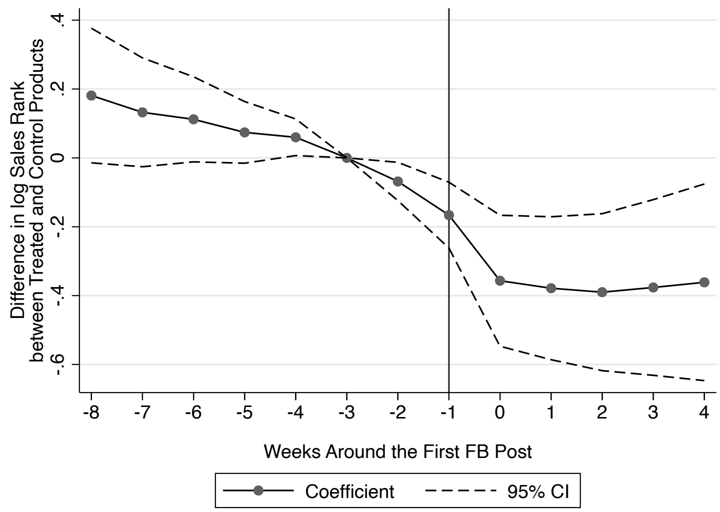
where everything is as in Equation 7, and $Week_{kit}$ represents a set of k dummies identifying 7-days intervals around the first Facebook post of each product. The λ_k coefficients can be interpreted as weekly treatment effects estimated before and after the treatment with respect to the baseline week -3.[18] We plot these estimates along with their 95% confidence intervals in Figure 13. Two findings emerge from this figure. First, while there is a decreasing trend in the pre-treatment period, the estimates before week -2 are indistinguishable from zero, suggesting that the parallel trends assumption is satisfied during this period. Second, there is a statistically significant increase in sales in weeks -1 and -2 relative to this baseline. This is consistent with the results in section 4.1 showing that for treated products sales begin to increase slightly early, suggesting that our DD analysis contains the same measurement error issue as the descriptive analysis. In our estimates of the size of the causal effects we measure the change in sales occurring after the first observed Facebook post (in week -1) and so to the extent that some of the increase in sales occurs before this, we may underestimate the size of the effect. Nevertheless, we do observe a large decrease in sales rank for treated products after week 0.

5.3 The effect of fake reviews on sales

To measure the causal effect of fake reviews on sales, we estimate Equation 7 using as the outcome the log of sales rank. We report these estimates in column 5 of Table 7. First, we find that the sales rank of control products increases about 22%. This is in line with the evidence we provided in Section 4.1 where we showed that products start recruiting

[18]We choose to set the baseline week to be -3 because, as we discussed in Section 4.1 we observe that for some products outcomes start to change at week -2.

Figure 13: The evolution of the treatment effect, i.e., the difference in log Sales Rank between treated and control products.



fake reviews when sales are falling. In the absence of fake reviews, sales are therefore likely to continue to fall and thus sales rank should increase. Second, and in line with what we observed in Figure 13, we estimate that compared to control products, treated products see a reduction in sales rank of 31%. The overall effect of fake reviews on sales rank for treated products ($\beta_1 + \beta_2$) is about 16%.

In Appendix D, we present several robustness checks that reinforce the causal interpretation of our results. First, we show that the sales estimates are not sensitive to the choice of the window around the mid-purge date used to select the set of control products. Second, to reduce concerns about our results being driven by the way in which we select control products, we consider a specification in which we define the treatment as a continuous variable rather than as a binary variable based on a time cutoff around the purge event. Third, we perform placebo tests where we re-estimate our results for fake purge dates, and find no difference in outcomes of treated and control products.

6 Evidence of Consumer Harm from Fake Reviews

We conclude the paper by evaluating whether consumers are harmed by fake reviews. To do so, we analyze the products' ratings after they stop buying fake reviews. If they continue receiving high ratings after rating manipulation ends it would be evidence that fake reviews are used by high-quality products in a manner akin to advertising. This would be consistent with the predictions of theoretical results in Dellarocas (2006) and others. If, by contrast, we see declining ratings and observe a large number of one-star reviews, it would suggest fake reviews are bought to mask low product quality and deceive consumers.

There is an inherent limitation in using ratings to infer welfare because consumers leave ratings for many reasons and generally ratings are not a literal expression of utility. But we argue that when products receive low ratings and a large number of one-star reviews, it indicates that the actual quality of these products is lower than what most customers expected at the time of their purchase. The low ratings are either a direct expression of product quality or an attempt to realign the average rating back toward the true level and away from the manipulated level. In this latter case, we still infer consumer harm, either because it indicates consumers paid a higher price than what they would have if the product was not overrated due to rating manipulation, or because the fake reviews 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 that fake reviews are a significant problem since they reflect negative consumer experiences that erodes trust in the platform's reputation system.[19]

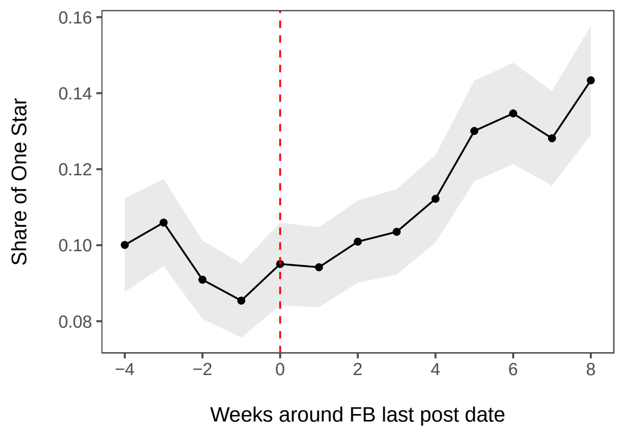
6.1 One-Star Ratings and Reviews

We previously showed in Figure 8 in Section 4.2 that average ratings fall after fake review recruiting ends. Figure 14 shows why. The share of one-star reviews increases by roughly

[19]Nosko and Tadelis (2015) show that when a buyer has a bad product experience with a third-party seller on a platform, they are significantly less likely to shop at that platform again.

70% after fake review recruiting stops. The increase in the share of one-star ratings and the increase in the total number of ratings mean that the absolute number of one-star reviews increases by even more.

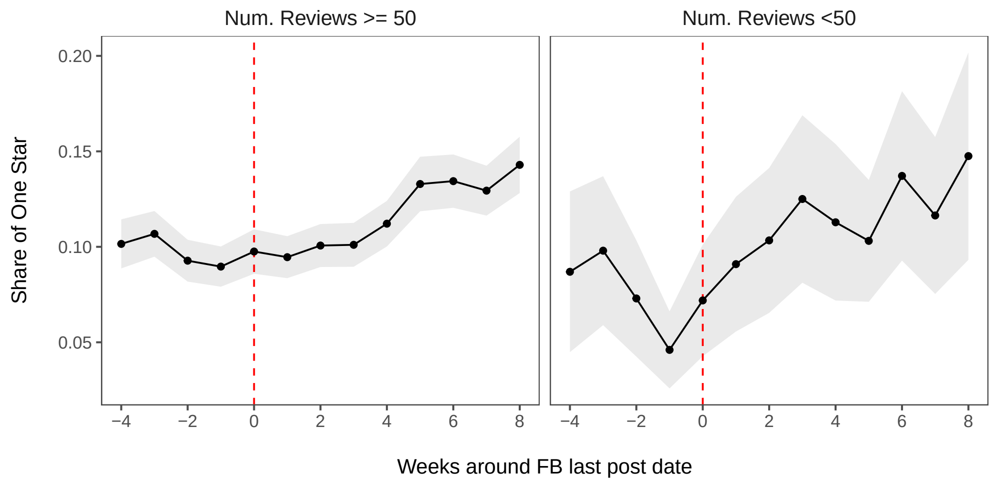
Figure 14: 7-day average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last time we observe Facebook fake review recruiting.



Next, we explore how this pattern varies for different types of products. It may be the case that ratings stay high for certain products. 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, which they can sustain if these products are high quality.

To test this, we segment products by number of reviews and age. Figure 15 shows how the share of one-star reviews changes for products with fewer than 50 reviews. The increase in one-star ratings is sharper for products with few reviews. Figure 16 makes the same comparison for products that have been listed on Amazon for fewer than 60 days. The young products experience a much larger increase in one-star reviews than the other products, with more than 20% of their ratings being one-star two months after they stop recruiting fake reviews. Overall, these results refute the idea that "cold-start" products use fake review efficiently. Instead, these products seem to be of especially low quality.

Figure 15: 7-day average share of one-star reviews before and after fake reviews recruiting stops by number of reviews accumulated prior to the fake review recruiting. The red dashed line indicates the last time we observe Facebook fake review recruiting.



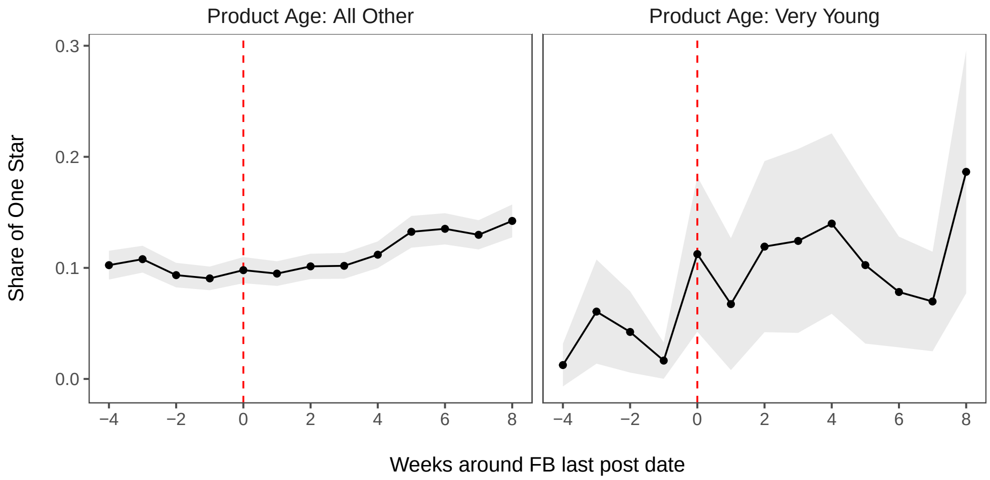
6.2 Text Analysis

So far, we have shown increases one-star reviews to provide evidence that consumers are harmed by rating manipulation. Here, we provide additional evidence by using state-of-the-art machine learning algorithms to analyze the text of these negative reviews.

The goal of this analysis is twofold. First, we want to test if the negative reviews posted after a product buys fake reviews are different from other negative reviews. It could be the case that one-star reviews increase after any sales spike and this is not a phenomenon specific to fake reviews. If so, test analysis should not be able to distinguish between them. Second, if they are indeed distinctive, we want to identify what text features differentiate them. Our simple model discussed in Section 3 shows that the returns to rating manipulation are higher for products with lower production costs, all else equal. It therefore predicts that negative reviews from these products are likely to focus on quality issues and value relative to price.

We perform two types of comparisons. First, we compare the post-campaign one-star reviews for fake review products to the one-star reviews for these same products prior to their first Facebook post. Second, we compare the post-campaign one-star reviews to one-star reviews for a different set of products that were not observed buying fake reviews.

Figure 16: 7-day average share of one-star reviews before and after fake reviews recruiting stops by product age (very young products are those listed for fewer than 60 days). The red dashed line indicates the last time we observe Facebook fake review recruiting.



We start by sampling 5,000 one-star reviews of each type: from products recruiting fake reviews prior to the first Facebook post, from those same products after the last Facebook post, and from a set of competitor products.[20] Then, we train a text-based classifier to predict whether each review is from either before or after fake review recruiting, or in the second test, from either a product recruiting fake reviews or not. Following standard practice, we split the review 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 ranges between 61% and 75% and a ROC-AUC score that varies between 66% and 83% for both types of comparisons. These results suggest that in both cases the classifier can distinguish between the different kinds of one-star reviews based on their text. In other words, even holding the products themselves and their star-rating constant, the reviews written for products after fake review recruiting contain significantly different text features compared with those written beforehand. Similarly, these reviews contain a significantly different set of words compared with

[20] As we discussed in Section 2, competitor products are defined as those products appearing on the same results page for a keyword search as the focal products.

reviews written for products that did not recruit fake reviews.

We next look at what are the most predictive text features for distinguishing the different product types. In Table 8, we compare the text features of negative reviews posted before and after rating manipulation by reporting the top 30 features. What emerges from this table is that one-star reviews written after rating manipulation occurs are predicted by text features mostly related to product quality ("work", "broke", "stop work") or value ("money", "waste money") or else explicitly suggest the consumer felt deceived or harmed ("return", "disappoint"). By contrast, the reviews for the same products prior to rating manipulation are associated with idiosyncratic product features, such as "earplug", "milk frother", or "duvet". Table 9 reports the top features for the model trained using fake review products and competitor products. Again, reviews for fake review products are associated with text features mostly related to product quality ("qualiti", "stop work", "work", etc.), value/price ("waste money", "money", "disappoint", etc.); instead, competitors' one-star reviews are predicted by text features mostly related to idiosyncratic product characteristic ("second attach", "fade", "reseal", etc.)

Overall, these results are consistent with one another and 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 8: Most Predictive Text Features: Before v After Fake Reviews

Period	Top 30 Text Features
Before recruiting fake reviews	muzzl, around neck, duvet, laundri, earplug, needless, milk frother, foam earplug, rectal, topper, espresso, lightn, like go, keep lick, nois reduct, degre differ, like tri, frizzi, espresso machin, wildli, breath, work never, expect much, concert, time open, stori, octob, inflat collar, unsaf, vinegar
After recruiting fake reviews	work, product, money, return, use, month, wast, time, would, wast money, stop, charg, like, even, disappoint, broke, stop work, week, first, tri, light, back, good, bought, batteri, qualiti, item, recommend, purchas, turn

Note: Model accuracy and ROC-AUC are 61% and 66%, respectively

Table 9: Most Predictive Text Features: Focal vs Non-Focal Products

Products	Top 30 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

Note: Model accuracy and ROC-AUC are 63% and 69%, respectively

7 Discussion and Conclusions

It has become commonplace for online sellers to manipulate their reputations on online platforms. In this paper, we study the market for fake Amazon product reviews, which takes place in private Facebook groups featuring millions of products. We find that soliciting reviews on Facebook is highly effective at improving several sellers' outcomes, such as number of reviews, ratings, search position rank, and sales rank. 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. Rating manipulation is not used efficiently by sellers to solve a cold-start problem, in other words.

We also find evidence that this practice is likely harmful to consumers, as fake review recruiters ultimate see a large decrease in ratings and increase in their share of one-star reviews. An important implication is that rating manipulation is also likely to harm honest sellers and the platform's reputation itself. If large numbers of low-quality sellers are using fake reviews, the signal value of high ratings could decrease, making consumers more skeptical of new, highly rated products. This, in turn, would make it harder for high-quality sellers to enter the market and would likely reduce innovation.

Firms are continuously improving and perfecting their manipulation strategies 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 manipulate their ratings 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.[21]

We also document that Amazon does delete large numbers of reviews and that these deletions are well-targeted, but there is a large lag before these reviews are deleted. The result is that this deletion policy does not eliminate the short-term profits from these reviews or the consumer harm they cause.

Of course, Amazon has other potential policy levers at its disposal to regulate fake reviews. But we do not observe Amazon deleting products or banning sellers as a result of them manipulating their ratings. Nor do we observe punishment in the products' organic ranking in keyword searches. This keyword ranking stays elevated several months after fake review recruiting has ended, even when Amazon finds and deletes many of the fake reviews posted on the platform. 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.

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

References

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

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

Chevalier, J. and Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1(2).

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*.

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.

He, S. and Hollenbeck, B. (2020). Sales and rank on amazon.com.

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.

Nosko, C. and Tadelis, S. (2015). The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment. *NBER Working Paper 20830*.

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

Rao, A. (2021). Deceptive claims using fake news marketing: The impact on consumers. volume Forthcoming.

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.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

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

Wilbur, K. and Zhu, Y. (2009). Click fraud. *Marketing Science*, 28(2):293–308.

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.

Appendix

A Sales Data

In this appendix, we first describe how we collect data on sales in units, and then how we convert sales rank to sales in units for instances in which this is unobserved. Amazon does not display metrics on sales quantities, only on an ordinal Best Seller Rank, a number that ranks products based on their rate of sales relative to other products in the same category.

To acquire sales quantity data, we exploit a feature of the Amazon website that allows us to infer the number of units of a product that are currently in stock. To observe a product's inventory, one must simply add to an Amazon cart increasing numbers of units of the product until the seller runs out of stock. At this point, Amazon will display an alert telling the buyer the total number of units available. The highest number of units that can be added to an Amazon cart is 999 and so for products with inventories below 1000 this method allows us to observe the number of units currently in stock. We employ research assistants to collect data using this method for a panel of products every 2 days. By observing inventories repeatedly over time, we can infer the rate of sales.

After collecting inventory data, we first remove observations in which the inventory is 0 or at the upper limit of 999 or if the seller has placed a limit on the number of units that can be purchased. We then calculate the difference in inventories between each two day period. We remove any observations where the inventory increases over this period. We use the remaining data to calculate sales per day. A more detailed description of this procedure and the resulting data can be found in He and Hollenbeck (2020). We observe data on sales in units for 683 of the focal products.

These data do not cover every period and, most importantly, we cannot observe sales data prior to the first Facebook post of these products. Therefore we estimate the relationship between sales rank and sales in units using the sales data to approximate the level of sales for these missing periods. To do so, we generalize the approach taken by Chevalier and Goolsbee (2003) and estimate a log-log regression with product fixed effects. This provides a good fit, with an adjusted-R^2 of .89. More details on the estimation and alternative models for estimated sales quantities are available in He and Hollenbeck (2020).

Lastly, we then use the regression estimates to infer the missing data on sales units at different dates for the same set of products based on their observed rank on those dates. We plot these outcomes in the short run and long run in Figures 5 and 9.

B Descriptive regression analysis

B.1 Short-term Analysis

We use data from the interval [-8,4] weeks around the first Facebook post and estimate the following equation on each outcome variable:

$$\mathrm{y}_{it} = \beta_1 \ \mathrm{After}_{it}^{\leq 2} + \beta_2 \ \mathrm{After}_{it}^{>2} + \alpha_i + \tau_t + \epsilon_{it}, \tag{9}$$

where $\mathrm{After}_{it}^{\leq 2}$ is a dummy for the time period from zero to two weeks after the beginning of the Facebook promotion and $\mathrm{After}_{it}^{>2}$ is a dummy for the time period after that. This divides up our sample into three periods: a before period, a period in which short-term changes should be present, and a period in which more persistent changes should be present. In each case we include year-week, τ_t, and product fixed effects, α_i. We include data on the 2,714 competitor products for which we have collected daily review data. These products are never observed buying fake reviews, so their After_{it} dummies are all set at zero.

The results for each variable for all products are shown in Table 10.[22] Consistent with our visual analysis, we see significant short-term increases in average rating, number of reviews, sales, and search position (keyword rank). The increase in weekly average rating is roughly .11 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 longer-term dummy for the number of reviews and search position, confirming that the changes in these variables are more persistent.

Next, we add interactions with an indicator for whether or not the product is new to Amazon with few reviews. New products without established reputations may have different

[22]The high R^2 are likely due to the inclusion of product and year-week fixed effects fixed effect.

Table 10: Short-term Outcomes After Recruiting Fake Reviews

	(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
$\leq$ 2 wks	0.107***	0.445***	-0.260***	-0.412***	0.044***	0.002	0.022***	0.022***	-0.013**
	(0.019)	(0.017)	(0.022)	(0.028)	(0.009)	(0.013)	(0.006)	(0.003)	(0.004)
> 2 wks	0.034	0.320***	-0.246***	-0.434***	0.061***	-0.007	0.003	0.018***	-0.016**
	(0.021)	(0.020)	(0.028)	(0.030)	(0.010)	(0.014)	(0.007)	(0.004)	(0.005)
N	186389	247218	193381	91733	94122	94122	186389	186389	92361
R^2	0.22	0.67	0.81	0.64	0.55	0.52	0.15	0.15	0.98

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\leq$ 2 wks	0.12***	0.44***	-0.24***	-0.41***	0.049***	0.0022	0.024***	0.016***	-0.013**
	(0.019)	(0.018)	(0.023)	(0.030)	(0.010)	(0.013)	(0.006)	(0.003)	(0.004)
$\leq$ 2 wks × Coldstart	-0.20**	0.091	-0.28**	-0.030	-0.078*	-0.0042	-0.033	0.078***	-0.0050
	(0.067)	(0.057)	(0.085)	(0.074)	(0.031)	(0.041)	(0.027)	(0.022)	(0.015)
> 2 wks	0.059**	0.31***	-0.22***	-0.44***	0.069***	-0.0051	0.0064	0.015***	-0.016**
	(0.022)	(0.021)	(0.029)	(0.031)	(0.011)	(0.014)	(0.007)	(0.004)	(0.005)
> 2 wks × Coldstart	-0.36***	0.14*	-0.35***	0.067	-0.11***	-0.021	-0.054	0.050*	0.0015
	(0.080)	(0.070)	(0.100)	(0.086)	(0.034)	(0.051)	(0.030)	(0.023)	(0.017)
N	186389	247218	193381	91733	94122	94122	186389	186389	92361
R^2	0.218	0.668	0.810	0.638	0.551	0.517	0.152	0.150	0.978

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.

incentives to buy fake reviews or different outcomes afterwards. We define these as "cold-start" products if they have been listed on Amazon for 4 or fewer months and have 8 or fewer reviews. This is roughly 10% of the observed products. We see in Table 10 that these products do have different outcomes, specifically that these products' sales increase by a much larger margin than for regular products. They also get a larger increase in number of reviews but do not see an increase in weekly average rating.[23]

[23] This last result is due to the fact that cold-start products frequently start out with a perfect five-star rating. When measured 2 weeks prior to their first Facebook post, we find that 83% of cold-start products have an average rating of 5.0 stars, leading to an overall average rating of 4.65 stars across products. This compares with an average rating of 4.35 stars for non-cold-start products. The high initial rating these products enjoy inevitably decreases as more reviews are added.

B.2 Long-term Regressions

Similar to how we presented results for the short-term outcomes, we now show the long-term results in a regression context. To do so, we take the interval [-4,8] weeks around the last Facebook post and regress each outcome variable on a dummy for the time period from one to three weeks afterward, as well as an additional dummy for the time period after that. In each case, we include year-week and product fixed effects. Results are shown in Table 11.

Table 11: Long-term Outcomes After Recruiting Fake Reviews

	(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
$\leq$ 2 wks	-0.033	0.060***	-0.052**	-0.17***	0.021***	-0.0028	-0.0078	0.0093**	-0.0073*
	(0.018)	(0.018)	(0.019)	(0.017)	(0.005)	(0.009)	(0.006)	(0.003)	(0.003)
> 2 wks	-0.16***	-0.24***	0.082**	-0.14***	0.036***	-0.0046	-0.043***	0.0028	-0.016***
	(0.020)	(0.020)	(0.027)	(0.021)	(0.007)	(0.010)	(0.006)	(0.003)	(0.004)
N	187640	249444	194840	97022	99409	99409	187640	187640	97543
R^2	0.219	0.668	0.811	0.647	0.559	0.518	0.146	0.147	0.979
$\leq$ 2 wks	-0.026	0.042*	-0.041*	-0.17***	0.022***	-0.0032	-0.0074	0.0069*	-0.0082**
	(0.019)	(0.018)	(0.020)	(0.018)	(0.006)	(0.009)	(0.006)	(0.003)	(0.003)
$\leq$ 2 wks × Coldstart	-0.12	0.26***	-0.15	0.044	-0.035	0.0025	-0.012	0.039	0.017
	(0.075)	(0.064)	(0.083)	(0.073)	(0.023)	(0.038)	(0.028)	(0.021)	(0.013)
> 2 wks	-0.15***	-0.25***	0.090**	-0.14***	0.040***	-0.0031	-0.040***	0.00070	-0.017***
	(0.020)	(0.021)	(0.027)	(0.021)	(0.007)	(0.011)	(0.006)	(0.003)	(0.004)
> 2 wks × Coldstart	-0.18*	0.19**	-0.11	0.11	-0.074*	-0.023	-0.047	0.038*	0.024
	(0.079)	(0.071)	(0.108)	(0.088)	(0.029)	(0.043)	(0.032)	(0.019)	(0.014)
N	187640	249444	194840	97022	99409	99409	187640	187640	97543
R^2	0.219	0.668	0.811	0.647	0.559	0.518	0.146	0.147	0.979

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.

The overall results, shown in the first two rows, confirm the visual analysis that shows that average ratings, number of reviews, sales and keyword position all fall after fake review recruiting ends. However, some of the increases in these variables are still present in the first week or two after the last Facebook post.

We also test interactions for "cold-start" products. We find that these products' ratings fall even further than for regular products, but that their increase in number of weekly reviews is more persistent. This is consistent with the fact that the decrease in sales rank is larger and more persistent for cold-start products. We don't find differences in terms of keyword rank, and find that the use of sponsored listings decreases for cold-start products while it increases for the rest of the products.

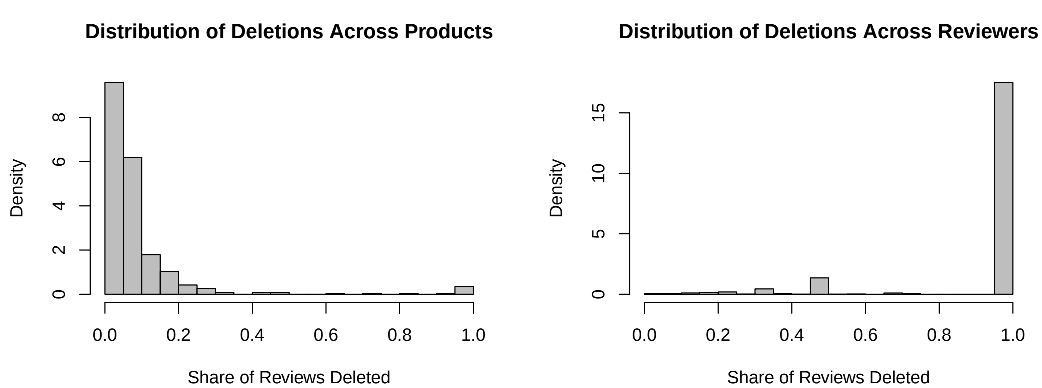
C Analysis of the mid-march Amazon purge

We have done an investigation of the patterns in review deletion across products, time, and reviewers in order to better understand the review "purge" and what selection criterion Amazon is using for these deletions. We focus first on the distribution of deletions across products and across reviewers to determine whether deletions are targeted at specific products buying fake reviews or at reviewers writing them. To give an example of this logic: if 10% of reviews were deleted during the review purge event, it could be that 10% of products were targeted and they all had 100% of their reviews deleted or it could be that specific products were not targeted and all products had about 10% of their reviews deleted. Similar analysis could find if individual reviewers were targeted or if the deletions are uniform across reviewers (of course these are extreme examples - reality must lie somewhere in between.)

We focus our investigation on the focal products (products observed buying fake reviews on Facebook) and find that during the 2-week period we call the review "purge", 3.2% of all 230,000 reviews are deleted. This is a small share of the total stock of reviews but in terms of the flow of deletions is many times higher than during normal periods. These deletions effect 40.6% of products (i.e. they have at least one review deleted) and 3.2% of reviewers.

This suggests deletions are targeted at a small group of specific reviewers and are not targeted at a narrow set of products. We next show the distribution of the share of reviews deleted at both the product and reviewer levels (conditional on having at least one review deleted.) We plot histograms of each in Figure 17.

Figure 17: Distribution of Deletions During Purge Event



This figure shows that the vast majority (93%) of products affected by the review deletion event have fewer than 20% of their reviews deleted and nearly half have fewer than 5% of their reviews deleted. Among reviewers, the opposite pattern holds. The vast majority (87.5%) of reviewers have all of their reviews deleted. This evidence is unfortunately biased, however, by the nature of our data collection. We initially only collect reviews at the daily basis for our focal products and so the set of reviewers we analyze here are those who have posted on these products in this time period. We did not scrape these reviewers' other reviews (for non-focal products) at the time, as would be required to track the full share of their reviews deleted at a given point in time. Therefore the vast majority (83%) of these reviewers have only 1 review observed to begin with.

Yet, among reviewers with more than 1 review who have reviews deleted, the same pattern does hold. In this group, reviewers with multiple reviews, at least one of which is deleted in the review purge, 77% have 100% of their reviews deleted. When we condition on reviewers having at least 5 reviews the share with all reviews deleted is 78%.

This analysis strongly suggests that individual products are not targeted when Amazon deleted large numbers of reviews in mid-March 2020 but rather that individual reviewers were targeted.

D DD Robustness checks

Sensitivity to the purge window Here we show that the sales estimates are not too sensitive to the choice of the window around the review purge used to select the set of control products. We do so by reporting in Table 12 the estimates for sales rank using three alternative windows around the mid-purge date: [-2,2] weeks, [-1,2] weeks, and [-1,1] weeks.

Continuous treatment To further reduce concerns about our results being driven by the way in which we select control products, here we show that our estimates are robust to a continuous definition of the treatment. To do so, for each product, we define a treatment variable, $\textit{log Purge Distance}_i$, which is equal to the log of the absolute value of the difference in days between the mid-purge date (March 15, 2020) and the date of the first Facebook post

Table 12: Diff-in-Diff using different purge windows

Purge Window	(1) [-2,2]	(2) [-1,2]	(3) [-1,1]
After	0.166*	0.178*	0.198
	(0.070)	(0.077)	(0.115)
After × Treated	-0.325***	-0.338***	-0.377**
	(0.086)	(0.092)	(0.131)
PSM Sample	Yes	Yes	Yes
N	12512	12512	11553
R^2	0.85	0.85	0.87

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.

of each product. We then estimate Equation 7, but replacing the binary treatment variable with this new continuous treatment. We report these results in Table 13. We observe that for small values of the treatment variable, i.e., for products whose first Facebook post is very close to the mid-purge date, there is a small and non statistically significant effect on reviews, and a positive effect on sales. However, the opposite is true for products whose first Facebook post is far from the mid-purge date.[24]

Table 13: Estimates using a continuous treatment variable

	(1) log Cum. Reviews	(2) Sponsored	(3) Coupon	(4) log Price	(5) log Sales Rank
After	0.040	-0.042	-0.034	-0.025	0.362*
	(0.070)	(0.047)	(0.067)	(0.019)	(0.146)
After × log Purge Distance	0.041*	0.019	0.009	0.004	-0.135***
	(0.019)	(0.013)	(0.018)	(0.005)	(0.037)
N	15789	9543	9543	9463	15077
R^2	0.93	0.64	0.67	0.99	0.87

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.

Placebo review purge To further reinforce the validity of our estimates, we perform a placebo test in which we create a placebo review purge by moving the mid-purge date either

[24]For example, at the median *log Purge Distance*$_i$ which is 3.89 (about 48 days), the increase in cumulative reviews is about 22% ($p < 0.01$) and the decrease is sales rank is about 15% ($p < 0.01$).

four weeks back or four weeks forward. We estimate Equation 7 using these thresholds and report these results in Table 14. As expected, we observe that recruiting fake reviews has a negative effect on sales rank for control products and that this effect is not different for treated products.[25]

Table 14: Estimates using placebo review purges

	(1) 4 weeks before	(2) 4 weeks after
After	-0.166*	-0.142*
	(0.079)	(0.060)
After × Treated	0.027	0.001
	(0.086)	(0.065)
N	15077	15077
R^2	0.87	0.87

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.

Alternative Propensity Score Matching algorithms Finally, we show that our results are not sensitive to the type of matching algorithm used. In Table 15 below, we report the estimates for sales rank using nearest-neighbor and local linear regression matching algorithms, an obtain results consistent with those reported in column 5 of Table 7.

[25] We do not use PSM in this exercise to further reinforce the fact that potential differences between treated and control products are not driving the sales effects reported in Table 7 (however, we obtain qualitatively similar results when we apply PSM). In addition, using the full data sample and the real purge, we obtain results consistent with those reported in Table 7.

Table 15: Estimates using alternative matching approaches

	(1) NN	(2) LLR
After	0.213*	0.188
	(0.100)	(0.098)
After × Treated	-0.368**	-0.370**
	(0.120)	(0.117)
N	7288	11489
R^2	0.88	0.87

Note: In column1 we report the results using the nearest-neighbor algorithm for matching with $n = 20$, and in column 2 we report the results using the local linear regression algorithm with a bandwidth of 0.005. 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.