Exploiting Social Media for Fake Reviews: Evidence from Amazon and Facebook

SHERRY HE
UCLA Anderson School of Management
and
BRETT HOLLENBECK
UCLA Anderson School of Management
and
DAVIDE PROSERPIO
USC Marshall School of Business

We provide an overview of our recent work that studies the market for fake product reviews on Amazon.com where reviews are purchased in large private internet groups on Facebook and other sites. 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 mostly used by high-quality or young products in a manner akin to advertising. 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 and is deceiving and harming consumers.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Economics, Measurement, Theory

Additional Key Words and Phrases: Fake reviews, eCommerce, online ratings

1. INTRODUCTION

Fake reviews and reputation manipulation tactics have become extremely commonplace on many online platforms. These practices are typically seen as harmful because they deceive consumers and erode trust in review platforms. There has consequently been a great deal of research interest in identifying, preventing, and mitigating fake reviews [Lappas et al. 2016; Luca and Zervas 2016; Mayzlin et al. 2014]. This prior research has almost entirely suffered from the drawback of not directly observing actual fake reviews or the resulting outcomes for those who seek them out. We set out to fill this gap by creating a large new dataset directly observing fake review buying and selling and tracking the outcomes affected by rating manipulation to better understand the economics of the fake review marketplace.

The market we study in [He et al. 2021] features Amazon sellers posting in private online groups to promote their products and solicit willing customers to purchase

Authors' addresses: `sherry.he.phd@anderson.ucla.edu`, `brett.hollenbeck@anderson.ucla.edu`, `davide.proserpio@marshall.usc.edu`

them and leave positive reviews in exchange for compensation. We collect data from this market by sending research assistants into these groups to document what products are buying fake reviews and the duration of these promotions. We then carefully track these products' outcomes on Amazon.com, including posted reviews, average ratings, prices, and sales rank. This is the first data of this kind, in that it provides direct evidence on both the fake reviews themselves and on detailed firm outcomes from buying fake reviews.

In general, because consumers value trustworthy information and e-commerce platforms value having good reputations, their incentives should be aligned in that they both want to avoid fake reviews. However, this may not always be the case. In particular, platforms may benefit from allowing fake positive reviews if these reviews increase their revenue by generating sales or allowing for higher prices. It may also be the case that fraudulent reviews are not misleading in the aggregate if higher quality firms are more likely to purchase them than lower quality firms [Dellarocas 2006]. They could be an efficient method for high-quality sellers to solve the "cold-start" problem and establish reputations. In this case, fake reviews may be seen as harmless substitutes for advertising rather than as malicious. It is therefore an empirical question whether firms and regulators should view rating manipulation as representing a significant threat to consumer welfare.

Our research objective is to answer a set of currently unsettled questions about online rating manipulation. How does this market work, in particular, what are the costs and benefits to sellers from buying fake reviews? What types of products buy fake reviews? 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?

Our sample consists of approximately $1,500$ products observed soliciting fake reviews over a nine-month period. We track the outcomes of these products before and after the buying of fake reviews using data collected from Amazon. The focus is on to what extent their ratings, reviews, and sales increase during this period. We also leverage a period in which Amazon mass deletes a large number of reviews to measure the causal effect of fake reviews on sales. We also track outcomes after the last observed post soliciting fake reviews to measure the long-run effects of fake reviews on sales and to observe whether organic ratings stay high or decline. This can help to provide evidence of consumer harm. If the products continue to receive high ratings from consumers after they stop buying reviews, it would suggest the fake reviews are more akin to advertising and are mainly bought by high quality products, potentially to solve a cold-start problem.

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

2. DESCRIPTIVE RESULTS ON PRODUCT OUTCOMES AFTER BUYING FAKE REVIEWS

We begin by quantifying the extent to which buying fake reviews is associated with short-term changes in average ratings, reviews, and sales rank. To evaluate these outcomes, we partition the time around the earliest Facebook recruiting post date (day 0) in 7-day intervals. We then plot the quantity of interest for eight 7-day intervals before fake reviews recruiting start and four 7-day intervals after fake reviews recruiting starts. We focus on roughly four weeks after fake reviews recruiting starts because initially we are interested in discussing short-term effects.

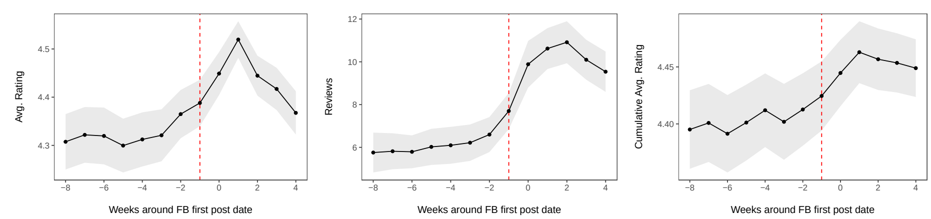

Fig. 1. 7-day average ratings, 7-day average number of reviews, and cumulative average ratings before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.

In the left panel of Figure 1 we plot the weekly average rating. It shows that the average ratings increases by about 5%, from 4.3 stars to 4.5 stars at its peak, after Amazon sellers start recruiting fake reviewers. This increase in rating is short-lived, and it starts dissipating just two weeks after the beginning of the recruiting of fake reviews; despite this, even four weeks after the beginning of the promotion, average ratings are still slightly higher than ratings in the pre-promotion period. In the middle panel of Figure 1, we plot the weekly average number of posted reviews. We observe that the number of reviews increases substantially around interval 0, nearly doubling, providing suggestive evidence that recruiting fake reviewers is effective at generating new product reviews at a fast pace. In the right panel of Figure 1, we plot the cumulative average rating before and after the Facebook promotion starts. We observe a positive change centered around the beginning of the promotion and that stabilized for about two weeks after the promotion begins, after which the increase starts to dissipate.

Next, we examine these products sales outcomes. In the left panel of Figure 2 we plot the average log of sales rank. It shows that recruiting fake reviewers is associated with a large and immediate decrease in sales rank (i.e., product sales increase). The center panel of Figure 2 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.[He and Hollenbeck 2020] describes how we collect this data and model the relationship. We plot the observed sales and point estimates of estimated sales around the time of the first FB post and see a sharp increase in average units sold, from around 16 units per week to roughly 20.

Another reason for higher sales is that products recruiting fake reviews will be ranked higher in the Amazon search results due to them having higher ratings and more reviews (both factors that are likely to play a role in determining a product search rank). To investigate whether this is the case, in the right panel of Figure 2 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.

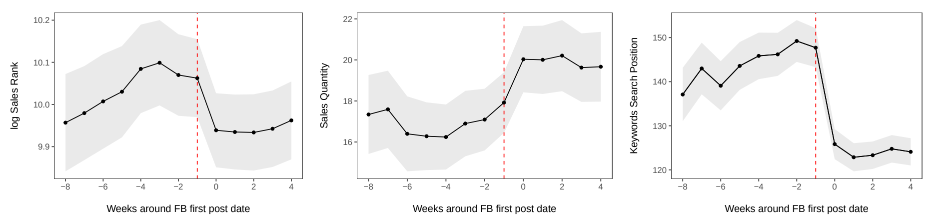

Fig. 2. 7-day average sales rank before and after fake reviews recruiting begins (left), sales in units (center), and keyword search position (right) before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.

3. MEASURING THE CAUSAL EFFECT OF FAKE REVIEWS ON SALES

The results presented so far are descriptive and should not be interpreted as measuring causal effects. There are two concerns in estimating the effect of rating manipulation on sales. The first is that sellers buying fake reviews may time these purchases around unobserved shocks to demand, either positive or negative. The second concern is that we observe that many sellers cut prices and increase advertising at the same time they recruit fake reviews, making it difficult to isolate the effect of fake reviews on sales.

Despite these issues, our data contains a temporary period during which Amazon instituted a sharp change in policy that allows us to isolate the causal effect of fake reviews, establish that this is a profitable strategy for sellers, and understand the magnitude of the effects that fake reviews can have on sales.

During mid-March of 2020 Amazon undertook a large-scale purge of reviews with much higher rates of deletion than normal. 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 use the products that recruited fake reviews just before or during the review purge as control products and all other products that recruited fake reviews at different times as treated products. We can then employ a difference-in-differences (DD) strategy that compares sales of treated products before and after they buy fake reviews with respect to a baseline of changes in sales of control products, and estimate the causal effect of rating manipulation on sales.

This 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 are able to establish empirically that each of these assumptions hold and so we can then estimate a standard DD regression which takes the following form:

$$y_{it} = \beta_1 \text{ Treated}_i + \beta_2 \text{ After}_{it} + \beta_3 \text{ Treated}_i \times \text{After}_{it} + \alpha_i + \tau_t + X_{it}^{'}\gamma + \epsilon_{it}, \quad (1)$$

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 that could be correlated with the outcome, and τ are year-week fixed effects to account for time-varying shocks to the outcome that affect all products (e.g., holidays).

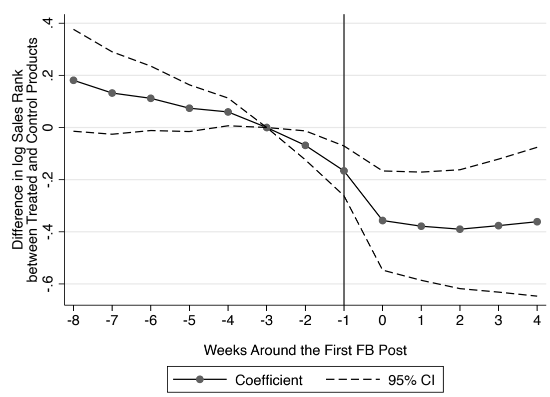

Fig. 3. The evolution of the treatment effect, i.e., the difference in log Sales Rank between treated and control products.

Before presenting the estimates of Equation 1, we visually present weekly changes of the difference in sales rank between treated and control products, before and after the treatment. We do so by estimating an even study regression in which we replace the interaction $\text{Treated}_i \times \text{After}_{it}$ in Equation 1 with treatment weekly leads and lags. We plot the resulting estimates of this regression, along with their 95% confidence intervals, in Figure 3. It shows that in the post-treatment period, there is a large decrease in sales rank for treated products associated with the start of the fake review recruiting (week 0), which points to a strong effect of fake reviews on sales. Using Equation 1, we estimate that the overall effect of fake reviews on sales rank for treated products is about 16%.

4. ARE CONSUMERS HARMED?

Next, we evaluate the potential harm to consumers from fake reviews. To do so, we analyze the long-term trends in consumer ratings for products observed being promoted using fake reviews. If these products continue to receive high ratings after the fake review recruiting period ends it would provide evidence that fake reviews are used by high-quality products in a manner akin to advertising. If, by contrast, we see declining ratings and observe a large number of one-star reviews, it may suggest that the sellers buying fake reviews are using them to mask the low quality of these products and deceive consumers into buying them.

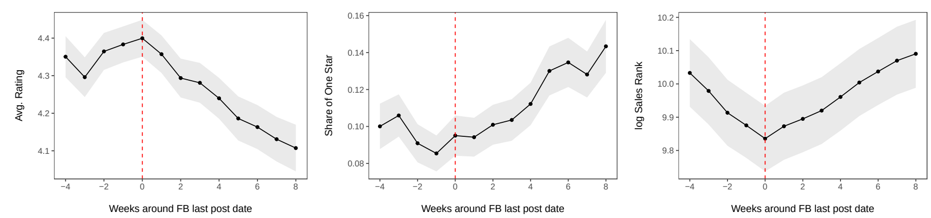

Fig. 4. 7-day average number of average ratings, reviews, and average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last week of data in which we observe Facebook fake review recruiting.

Figure 4 shows the outcomes for fake review products measured relative to the date of their last observed Facebook post. The red line thus indicates the date at which fake review recruiting stops. In the left panel, we show the average weekly star rating. It shows that average ratings decline quickly after rating manipulation stops, falling from 4.4 stars to 4.1 stars eight weeks later. The center panel shows results for the share of one-star reviews. This increases substantially and explains why the average rating is falling. Finally, the right panel shows results for the log of sales rank. It shows that after rating manipulation ends, sales decline significantly (shown as an increase in sales rank.)

These results suggest that fake review recruiting does not lead to a self-sustaining boost in sales and positive organic reviews. Instead, it leads to a backlash effect with large numbers of one-star reviews and declining sales.

5. DISCUSSION AND CONCLUSIONS

It has become commonplace for online sellers to manipulate their reputations on online platforms. Our research finds that rating manipulation is an effective strategy for generating a substantial boost in ratings, keyword position, and ultimately sales. Firms therefore have a strong incentive to continuously improve and perfect their manipulation strategies to avoid detection.

At the same time, our results suggest that, in contrast to some economic theory, these practices are not primarily used by high-quality sellers along the same lines as advertising. Instead, the long-term trends in ratings suggest these are primarily low-quality products and consumers are being harmed by the manipulated ratings.

Together, these results show why studying and understanding how firms manipulate their ratings, and how platforms can detect and remove them, continue to be extremely important topics of research for both academics and practitioners.

REFERENCES

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

HE, S. AND HOLLENBECK, B. 2020. Sales and rank on amazon.com. *Technical Note*.

HE, S., HOLLENBECK, B., AND PROSERPIO, D. 2021. The market for fake reviews. In *The 22nd ACM Conference on Economics and Computation (EC)*, P. Biró, S. Chawla, and F. Echenique, Eds. ACM, 588.

LAPPAS, T., SABNIS, G., AND VALKANAS, G. 2016. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research 27,* 4, 940–961.

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.