Informative Advertising or Deception? An Investigation of Review Fraud on Amazon

Maxwell C. Pilloff

Submitted to the Distinguished Majors Program
Department of Economics
University of Virginia
April 2023
Advisor: Federico Ciliberto

Informative Advertising or Deception? An Investigation of Review Fraud on Amazon

Maxwell C. Pilloff*

Abstract

User-generated reviews have become a crucial aspect of online shopping, and as review platforms have grown in popularity, review fraud has increased in prevalence. While this conduct is typically associated with low-quality firms that aim to deceive customers, this paper explored the possibility that fake reviews could be purchased by high-quality new products on Amazon to signal their quality to consumers and distinguish themselves from inferior products. A novel theoretical framework using a Bayesian approach was proposed to investigate the impact of review fraud on consumer learning and Monte Carlo simulations were run to evaluate the number of reviews necessary for customers to distinguish between a high and low-quality product. Model parameters and possible amounts of fraud were then estimated using data that tracks Amazon products that purchase fake reviews, but the findings were inconclusive. Investigating further, a series of OLS regressions were conducted, which provided evidence that higher perceived quality negatively predicts the estimated amount of review fraud. This suggests that producers of low-quality goods are more likely to buy a greater number of fake reviews. The results of this paper provide a warning to consumers about the presence of fraudulent reviews and motivation for the FTC and eCommerce sites such as Amazon to more vigilantly police review fraud.

*Completing this thesis was one of the most challenging things I've ever done. First, I would like to thank Professor Federico Ciliberto, whose advice and ideas guided me throughout this process. I am also grateful for Professor Sarah Turner, who supported me and helped me work through research questions even when this undertaking felt impossible. Finally, this paper would not have been possible without the support of my friends and family. Dad, thank you for sparking my love and interest in Economics from a young age; Mom, thank you for teaching me what hard work and dedication look like.

Table of Contents

1 Introduction

Studies have shown that 82% of American adults read customer reviews before purchasing a product for the first time (Smith & Anderson, 2016). While reviews may be expected to alleviate the issue of asymmetric information in eCommerce, numerous studies have discovered pervasive online review manipulation over the past decade. This involves a firm writing a positive review for itself or paying someone else to do so; it also occurs when a company creates negative reviews to harm close competitors.

Illegal review suppression and manipulation has been under increased scrutiny from the Federal Trade Commission (FTC), which has recently punished several firms for this activity. Samuel Levine, the Director of the FTC's Bureau of Consumer Protection, states that although "companies should know by now that fake reviews are illegal, this scourge persists" ("FTC to Explore Rulemaking to Combat Fake Reviews and Other Deceptive Endorsements", 2022). The FTC affirms that the current eCommerce landscape has made it easier for some firms to use fake reviews to promote their product or disparage a competitor, behavior that is commonly associated with low-quality products and can help boost sales. Additionally, it can be difficult for consumers, platforms, and competitors to distinguish fake reviews from organic ones ("FTC to Explore Rulemaking to Combat Fake Reviews and Other Deceptive Endorsements", 2022).

However, I posit that review fraud could also serve as informative advertising for high-quality products that are relatively unknown in the marketplace and wish to signal their quality to consumers. The number of sales and subsequent reviews necessary for a customer to determine a product's true quality would therefore decrease, leading to possible consumer welfare gains due to the manipulation.

The objective of this paper is threefold: to build a theoretical framework analyzing the number of reviews required for consumers to learn about a product's quality, to use the model to approximate the amount of review fraud on Amazon, and to empirically test the effect of quality on the number of fake reviews purchased.

First, I create a novel conceptual framework using a Bayesian learning approach in which firms of high and low-quality purchase fake reviews for a new product on Amazon. Using Monte Carlo simulations, I evaluate the number of reviews required for customers to distinguish between different quality products. Next, I estimate the prevalence of review fraud on Amazon using this model. I determine whether high or low-quality firms purchase more fake reviews and provide plausible estimates of review fraud. Finally, I check the robustness of the outcomes of the model and further examine the effect of product quality on the number of fake reviews purchased by a firm by running a series of regressions.

To my knowledge, review fraud has sparingly been analyzed or empirically investigated as a means of advertising among firms of different qualities. A handful of studies formulate conceptual frameworks that examine the impact of fake reviews on consumer welfare, but these models are not tested using data. Additionally, few papers have provided substantial evidence to support the FTC's claim that low-quality firms typically commit review fraud.

The content and results of this paper are of interest to many entities. They can help direct the FTC and other governing agencies in their enforcement of review manipulation. Also, online marketplaces such as Amazon can administer more efficient policies; if high-quality firms commit more review fraud, it could be a waste of resources for Amazon to identify and remove every fake review. Finally, the information can help inform consumers to make more thoughtful purchasing decisions on online marketplaces.

The structure of this paper is as follows: Section 2 reviews recent literature on review manipulation, quality signaling, and Bayesian learning models. Section 3 details the data and the collection process and provides some basic descriptive statistics as well. In Section 4, I develop a theoretical framework to determine the number of reviews needed to differentiate a high and low-quality product without fake reviews and empirically estimate model parameters. Section 5 alters this framework and allows a firm to buy fake reviews. I use data that tracks Amazon products that purchase fake reviews to investigate the number of fake reviews purchased by high and low-quality firms. Section 6 tests the model's results using a different empirical framework, and Section 7 discusses the limitations of the paper and ideas for future work.

The findings vary depending on the methodology. As outlined in Sections 4 and 5, slightly more reviews are needed to reveal the quality of products that purchase fake reviews. According to the model proposed in Section 5, this does not provide enough evidence to conclude that low-quality firms purchase more fake reviews. However, in Section 6, the estimated product quality is a statistically significant regressor that negatively affects the predicted amount of review fraud.

2 Literature Review

I explore literature across a few different topic areas, including theoretical and empirical investigations of online review fraud, quality signaling through advertising, and economic models of learning using Bayesian inference. To my knowledge, no prior research has combined these three disciplines. The study of review fraud is novel, with few papers aimed at developing a canonical model illustrating the effect of review fraud on consumers, other online sellers, and eCommerce platforms. Existing research primarily attempts to empirically detect fake reviews and estimate their prevalence depending on market structure and firm characteristics. Conversely, quality signaling models have been extensively researched in advertising theory, while Bayesian learning models are commonly used in data science and other academic fields.

2.1 Online Review Manipulation

Within advertising literature, research primarily examines the effects of truthful advertisements (e.g. Anderson & Renault, 2006; Bagwell, 2007), including truthful online reviews.

However, since the late 2000s, the theoretical analysis of false or manipulative reviews has grown. Several articles focus on the theoretical impact of review fraud on firm profits and consumer surplus but vary in their approach. Some models, such as the one presented by Hattori and Higashida (2012), suggest that in certain inefficient market conditions, the rise in output stemming from deceptive advertising can increase social welfare. On the other hand, Piccolo et al. (2015) conclude that there are settings in which review fraud can benefit consumers; even though a consumer may get tricked by a low-quality firm, there is a resulting downward pressure on prices due to subsequent truthful negative reviews that increases social welfare.

Another contribution to the literature is Rhodes and Wilson (2018), who propose a model that considers an eCommerce site's ability to penalize companies that engage in review manipulation. The authors discover that it may be optimal for the platform to induce false advertising for firms with moderate product quality. Additionally, in certain market conditions, fraudulent reviews can increase social welfare. To my knowledge, this is among the few papers considering fake reviews as a form of advertisement, and it directs my theoretical and empirical work.

However, not all models conclude that review fraud can increase consumer surplus. Glazer et al. (2018) investigate a consumer's learning process and contend that counterfeit reviews are always unproductive. However, they do not consider how product quality can influence this learning process.

The empirical effects of review fraud have also been widely studied. Mayzlin et al. (2014) investigate the presence of review manipulation in the hotel industry using a difference in the policies of two review platforms. They find that the incidence of fraud is higher for small, independent firms. Additionally, negative review manipulation against competitors increases when firms are small and independent. Luca and Zervas (2016) find comparable results in the restaurant industry.

Furthermore, Hollenbeck et al. (2019) examine how hotels change their ad spending in the presence of online reviews and find a statistically significant decrease in ad spending as online reviews become more positive, which suggests that ad spend and online review reputation can be seen as substitutes. This supports the FTC's assertion that low-quality firms are more likely to engage in review fraud, and they discover that this phenomenon is even more pronounced for smaller, independent hotel chains.

Finally, He et al. (2022) investigate the impact of fake reviews on Amazon and discover that rating manipulation leads to a causal increase in the short-term number of sales, as well as an increase in the average rating and number of reviews. As detailed in Section 3, I utilize the data collected in this study, which identifies specific firms that purchase fake reviews. The objective of this paper is to integrate theoretical work with recent advances in data availability.

2.2 Advertising Theory

Quality signaling is the idea that an agent can implicitly communicate information about itself to another party. Milgrom and Roberts (1986) apply this concept to advertising theory, where the firm selects the number of resources it wishes to allocate to advertising, and this indirectly communicates the product's quality to consumers. Kihlstrom and Riordan (1984) also create a similar model. They find that in a repeated game, there exists a separating equilibrium; the high-quality firm can signal its superiority by investing a sufficiently high amount in advertising during the first period. Although it may incur initial losses, the firm can recoup these losses in future periods. The low-quality firm, on the other hand, cannot afford to spend such a high amount on advertising as it cannot recoup potential losses in future periods.

2.3 Learning Models

Finally, I explore various models of economic learning. The Bayesian learning framework is structured as follows: an agent has a prior distribution, which summarizes their best initial estimate of a global parameter. Then, the agent receives signals and extracts information, which they use to update their prior distribution and create a posterior distribution. This is a revised estimate of the global parameter given the prior distribution and the new information

(Kelly & Kolstad, 1999). I focus on passive Bayesian learning models, where agents do not actively seek out information and signals are exogenous. In Sections 4 and 5, consumers on eCommerce sites are agents, and online reviews serve as exogenous signals.

Learning models are common in economic literature and have different applications; some determine a single agent's utility-maximizing decisions (e.g Rothschild, 1974), while others involve many agents and explore games in which an individual's actions provide information to others (e.g Banerjee & Fudenberg, 2004). Typically, these models involve passive information flow and explain social learning. Additionally, in these models, given that the environment is stationary and information is free to acquire, the agent is able to learn and make optimal decisions (Sobel, 2000).

The structure of this paper mirrors that of other studies which create economic models of learning (e.g. Lee & Moretti, 2009). First, I explain and draw conclusions from a model and then test the robustness of these results.

3 Data

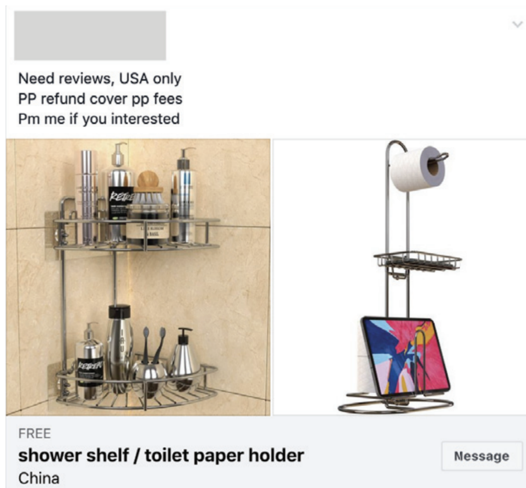
The data used in this paper was obtained from He et al. (2022). In this section, I provide a detailed overview of the available data and the collection process, followed by descriptive statistics.

3.1 Facebook Groups and Data Collection

Facebook has become a popular platform for sellers to buy fake reviews for their Amazon products. To do so, they post a request in private groups consisting of members who are willing to publish a five-star review for any product and communicate with all potential reviewers through Facebook private messages. Typically, reviewers receive a full refund for the product as well as an additional commission via PayPal after the fake review has been posted.

These reviews are often indistinguishable from genuine reviews and appear organic. Since reviewers have purchased the product on Amazon, their reviews are classified as a "Verified Purchase." Furthermore, since they are paid after their five-star reviews have been posted, reviewers are motivated to post something that is both well-written and positive. Figure 1 illustrates an example of one such Facebook post.

Figure 1: Example Facebook Post Seeking Fake Reviews (He et al., 2022)



From October 2019 to June 2020, the researchers monitored the activity of the 30 most popular of these Facebook groups. Each group had an average of roughly 16,000 members and over 550 daily requests. To gather data, the research team quasi-randomly selected posts in these Facebook groups, attempting to ignore the product type and characteristics of the post, and obtained the Amazon URL of the product using the keywords provided by the seller.

As a result of this process, 1492 products were identified. The following information was recorded for each product: the search keywords, product ID, product category and subcategory (from the Amazon product page), date of the Facebook post, the earliest post date for the same product, and the Facebook group name.

3.2 Amazon and Data Collection

The researchers gathered weekly information on Amazon review activity for each product that was identified. They recorded the number of weekly new reviews, the average rating of the weekly new reviews, the weekly share of one-star reviews, the total number of reviews, and the cumulative average rating. Qualitative measures such as review text, the presence of photos, and helpful votes were also collected.

Additionally, the researchers assembled a "control" group. For each product that was observed purchasing fake reviews, they found the two competitor products that appeared most frequently on the Amazon search page in the week before and after a product's first post on Facebook. This methodology allowed for the creation of a control group of products that closely resembled the focal group of products that were discovered to be purchasing fake reviews. The same review statistics were collected for these products.

Finally, weekly review activity was matched with the Facebook data by the number of weeks before or after the first and final Facebook post was observed (e.g., one observation could be four weeks before the first post and another could be 13 weeks after).

3.3 Statistics on Product Details

He et al. (2022) provide an overview of the group of products that were caught purchasing reviews. Table 1 calculates the number of items in each category and subcategory, as denoted by Amazon. The most frequently identified categories are "Beauty & Personal Care," "Health & Household," and "Home & Kitchen." The product set covers a wide range of industries, indicating that review fraud occurs for many types of goods on Amazon.

Table 1: Top Product Categories and Subcategories (He et al., 2022)

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	Outdoor String Lights	9

I identify 796 products in the focal group and 1336 products in the control group as new to the marketplace, with their first review recorded at the beginning of data collection. Among these products, 28 in the focal group and 447 in the control group did not have an average of $1, 2, 3, 4$, or 5 stars following the first review, indicating a possible error in data collection. A significant portion of the products in this dataset are new to the market, and hence satisfy the model conditions in Sections 4 and 5.

I analyze the review activity among both falsely reviewed products and the competitor group when data collection ended, as shown in Table 2. Products in the focal group have a higher average star rating than those in the control group. Additionally, they have roughly the same number of reviews, higher prices, and are more popular based on Amazon's sales rank, a method that sorts products in the same category by popularity. The lower the sales rank, the higher the sales of a product.

Table 2: Amazon Characteristics of Products in the Focal and Control Groups

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Average Rating					
Focal Group	4.33	0.52	4.08	4.45	4.69
Control Group	4.01	0.61	3.69	4.13	4.45
Number of Reviews					
Focal Group	253.07	529.15	33.00	81.00	243.00
Control Group	254.64	594.17	27.00	76.00	207.00
Selling Price					
Focal Group	33.52	67.13	13.99	19.99	32.99
Control Group	26.42	31.17	11.61	17.99	25.74
Sales Rank					
Focal Group	110677.82	162977.82	13653.27	43461.65	138045.75
Control Group	236513.75	225838.91	67405.52	145209.40	336453.35

He et al. (2022) also examine the country of origin for products in the focal group, matching the sellers' names on Amazon with records from the U.S. Trademark Office. They find that 84% of the sellers are based in mainland China, while 12% are from the United States.

3.4 Length of Facebook Posting

Statistics on the number of fake reviews purchased by each product are limited since it is challenging to differentiate between organic and fraudulent reviews. However, He et al. (2022) have recorded the number in full weeks since a firm first and last requested fake reviews on Facebook. Therefore, I calculate the number of weeks that a firm seeks reviews on Facebook (Table 3). The majority of products post on Facebook for fewer than two full weeks; one possible explanation for this short time frame is that firms that post on Facebook for longer are more likely to be caught by a government agency or an eCommerce site such as Amazon. Additionally, 226 firms post for just one week, while 570 post multiple times over multiple weeks.

Table 3: Summary Statistics of Facebook Posting Length

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Weeks Posting on Facebook	4.76	5.62	1	2	6

4 Model without Review Fraud

In this section, I present a model that demonstrates and estimates how many reviews consumers need to distinguish between a low and a high-quality new product on Amazon, and I estimate its parameters using the competitor group of products.

4.1 Assumptions

I first explain the fundamental assumptions of the model. All consumers are assumed to be homogeneous in terms of their likelihood to enjoy a product. The product quality $\theta \in \{h, \ell\}$ is either high (h) or low (ℓ). A high-quality product has a probability $p_h \in [0,1]$ to be enjoyed by a consumer, while a low-quality product has a probability $p_\ell \in [0,1]$ to be enjoyed with $p_h > p_\ell$.

Additionally, sales are exogenous in the model and a fixed proportion of sales leave a review. Therefore, for a fixed $\alpha \in (0,1]$, following q sales exactly αq reviews are posted. Each review indicates whether or not the product was enjoyed by the consumer and provides no further information. It is also assumed that the proportion of high and low-quality products in the population is equal, and consumers consider this when forming a prior distribution. Lastly, in this model, firms are not allowed to purchase fake reviews.

4.2 Model Creation

Let X_h and X_ℓ be random variables that record whether a reviewer enjoyed a product with quality h and ℓ respectively. Since the product is assumed to be new to Amazon, consumers have no prior knowledge of the product's quality and would find it challenging to estimate p_h. Given that the proportion of high and low-quality products in the population is equal, consumers believe that there is an equal probability of $p_h \in [0,1]$, and thus the prior distribution of p_h follows a Beta$(1,1)$ distribution.

Additionally, if αq reviews are posted, each review is randomly sampled with a probability

p_h to be positive and a probability $1 - p_h$ to be negative. Thus, $X_h|p_h \sim \text{Bin}(\alpha q, p_h)$. Consumers read these reviews and update their beliefs about the parameter value of p_h. As proved in the appendix, the posterior distribution for p_h is $p(p_h|X_h) \sim \text{Beta}(X_h+1, \alpha q - X_h + 1)$, where X_h is the number of positive reviews received. This process occurs symmetrically for a low-quality product.

4.2.1 Estimating p_h

I estimate p_h, the global proportion of consumers who enjoy a high-quality product on Amazon by analyzing firm behavior. Because the star average shown on Amazon is rounded to the nearest half-star, firms may manipulate their ratings to appear as higher-quality products (Luca, 2011). For example, a product with an average star rating of 4.24 would display 4 stars, while one with an average rating of 4.26 would display 4.5.

To investigate whether firms manipulate their ratings to surpass the 4.25-star threshold and appear as higher-quality products, I conduct a McCrary Sorting test (McCrary, 2008) as per Hollenbeck et al. (2019). This test examines the existence of a discontinuity in the density function around the 4.25-star cutoff. In the absence of manipulation, I would expect the density of average ratings to be continuous. However, the null hypothesis of the McCrary Sorting test is rejected ($p = .008$), so there is a "jump" in the number of firms that have cumulative averages above this cutoff (Figure A.1). Firms recognize that it is advantageous to raise their average rating above 4.25 stars and display 4.5 stars on Amazon.

Therefore, I predict that consumers perceive products with an average of 4.5 stars to be "high-quality." Since the model assumes reviews are either positive or negative, maintaining an average rating of 4.5 stars requires 87.5% of posted reviews to be positive if reviews are only 1 or 5 stars. So, I estimate $p_h = .875$.

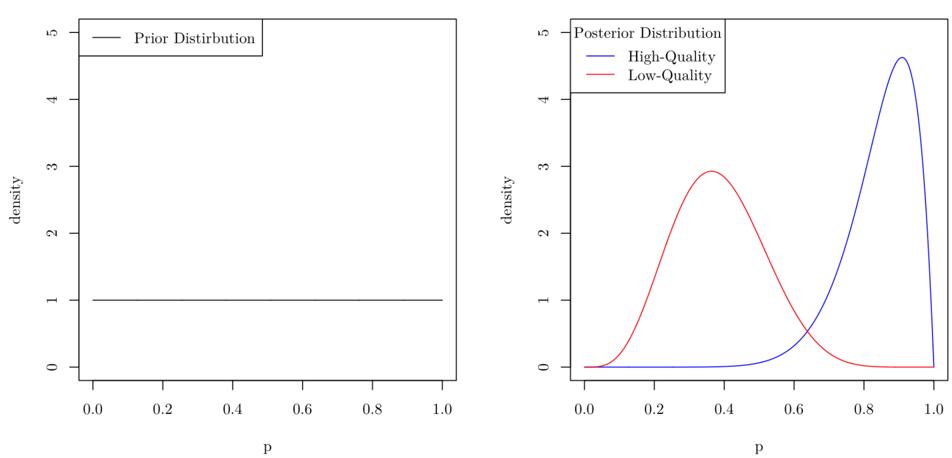
4.3 Monte Carlo Simulations

To determine the number of reviews needed to differentiate between a good and bad product, I conduct repeated Monte Carlo simulations using the model above. Because $X_h|p_h$ and $X_\ell|p_\ell$ are randomly sampled, I can estimate the distribution for the number of reviews by running 1000 repeated trials.

In each trial, I begin with two products of differing quality (h and ℓ) without any reviews. Then, I fix $p_h = .875$ and p_ℓ and sequentially "post reviews" for each product by randomly sampling from a Bernoulli distribution with parameters p_h and p_ℓ respectively. Following each review, I compare the posterior distributions and assume that quality is revealed when the posterior distributions are significantly different. To determine this, I perform a Kolmogorov–Smirnov test on a random sample of four points from each distribution. Reviews are simulated and continually added until the Kolmogorov–Smirnov test is significant at the 5% significance level.

For example, suppose $p_\ell = .5$. In one trial, 11 reviews are required for the posterior distributions to be different from each other. Figure 2 shows the prior and posterior distributions following 11 reviews in this example.

Figure 2: Example Prior and Posterior Distributions when Quality is Revealed

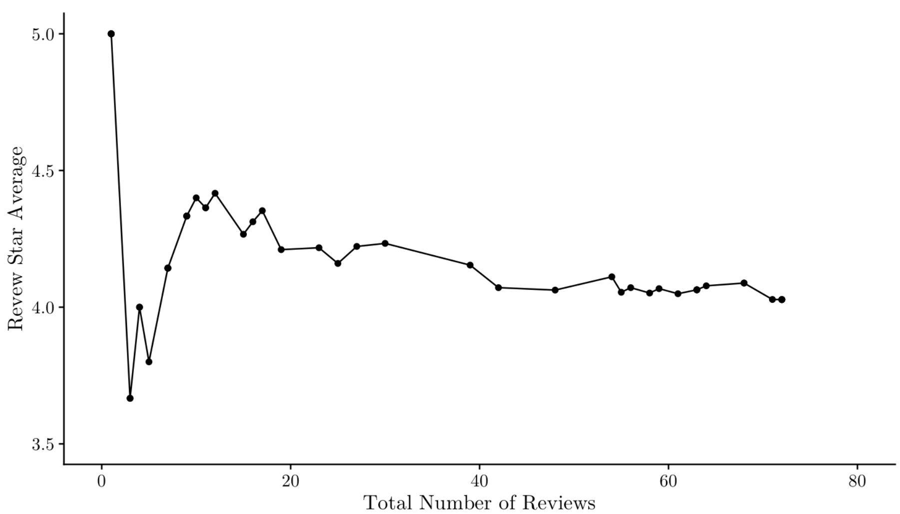


Since p_ℓ is unknown, I fix various values for p_ℓ and compute the mean and median of the simulated distributions (Figure A.2). As anticipated, when p_ℓ increases and there is a decrease in the difference between a high and low-quality good, more reviews are needed to reveal quality. For the lowest possible value of p_ℓ that I considered ($p_\ell = 0.5$), the mean number of simulated reviews to reveal quality is 10, and the median is 8; for the highest possible value ($p_\ell = 0.75$), the mean is 27 and the median is 19. The mean of each distribution is greater than the median, indicating that they are right-skewed.

4.4 Estimating p_ℓ

I use the control group of products and the model above to estimate p_ℓ. To do so, I must first determine the number of reviews necessary to reveal product quality in the data. Figure 3 illustrates the change in cumulative average rating as more reviews are posted for a product.

Figure 3: Cumulative Average Star Rating and Number of Reviews (ID = 2911)



As more reviews for the good are posted, more information becomes available. Eventually, the average rating of the product begins to converge to a certain value, as seen in Figure 3, which I conjecture represents its true quality. To determine the number of reviews needed to reveal product quality, I estimate how many reviews are typically needed for this convergence, which occurs when the change in the cumulative average rating becomes very small following a new review.

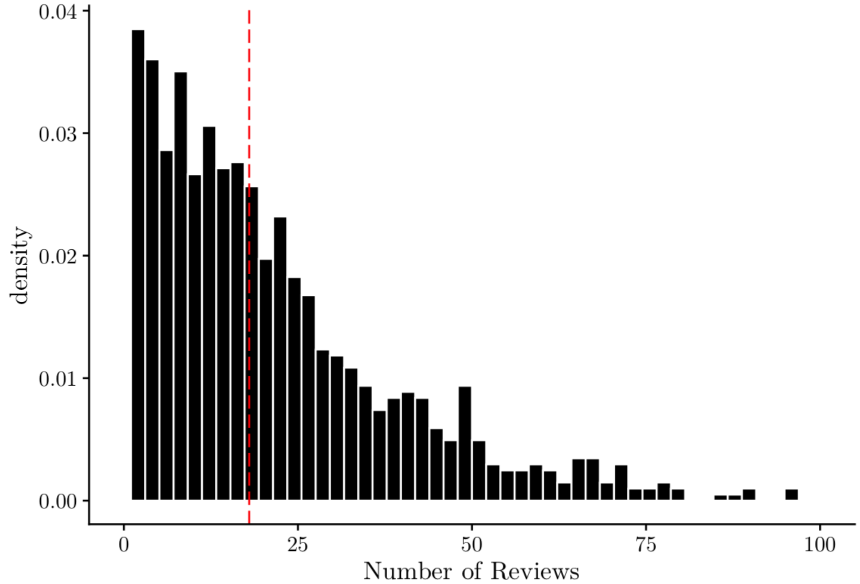
However, since reviews are compiled on a weekly basis, I am unable to observe the impact of each individual review on the cumulative average. Therefore, I calculate the slope as follows to find the average per review change in the average rating, where $cumavg_i$ is the cumulative average rating in week i and $numreviews_i$ is the total number of reviews in that week.

$$\frac{|cumavg_i - cumavg_{i-1}|}{numreviews_i - numreviews_{i-1}} \tag{1}$$

I use an absolute change of .025 as the necessary cutoff to determine whether a new review significantly impacts the average rating displayed on Amazon since there is a high probability the displayed rating does not change following an increase or decrease by this amount. Figure A.3 shows the average absolute per-review change in the average star rating as more reviews are posted for every product in the control group.

Through a one-sample Wilcoxon signed rank test, I identify the minimum number of reviews required for the median change among all products to be statistically significantly less than .025. When 20 reviews are posted, the null hypothesis that the median is equal to .025 is rejected ($p < .05$), so I conclude that this is the number of reviews needed to differentiate between high and low-quality products. From Figure 4 and Table A.1, if $p_\ell = .725$, then the median of the simulated distribution is 18 and the mean is 24.11. So, I estimate that $p_h = .875$ and $p_\ell = .725$.

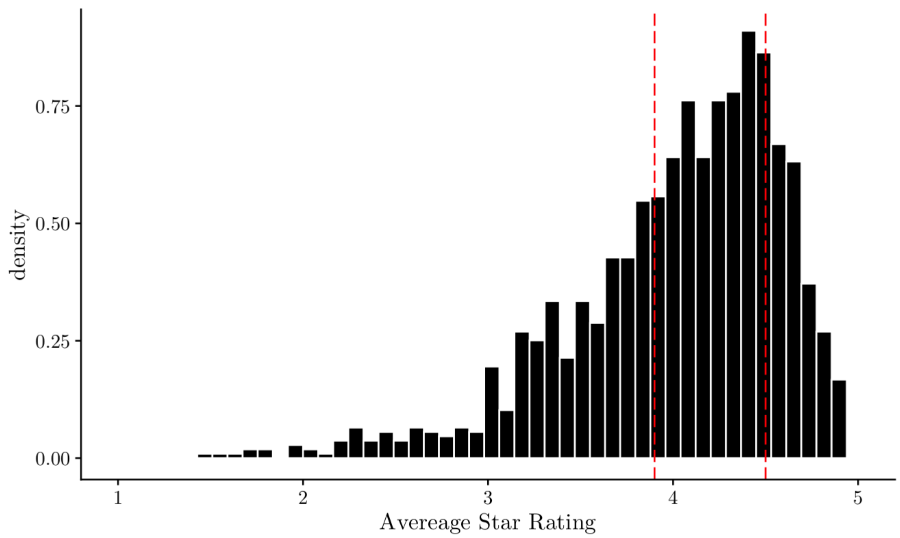
Figure 4: Simulated Distribution of Quality Revealing Reviews ($p_h = .875$ and $p_\ell = .725$)



4.5 Analysis of Parameter Estimation

Before considering the impact of product quality on buying fake reviews, I assess the validity of the estimates for p_h and p_ℓ. Figure 5 displays a histogram of the cumulative average ratings for all products in the control group when data collection ended. Drop lines implied by the estimates for p_h and p_ℓ are included. If only one and five-star reviews are posted, the low-quality product ($p_\ell = .725$) would have an average star rating of 3.9, and the high-quality product ($p_h = .875$) would have an average of 4.5.

Figure 5: Histogram of Average Star Rating for Control Group



The estimates for p_h and p_ℓ appear reasonable. Most products have an average rating between 3.5 and 5 stars, making them easily categorized as either high or low-quality products. Only goods with an extremely low average rating cannot be classified based on the estimated value of p_h and p_ℓ.

5 Model with Review Fraud

In this section, I build on the framework created in Section 4, but firms now have the ability to purchase fake reviews before beginning to sell their product.

5.1 Model Creation and Assumptions

The assumptions of this model are similar to those in the previous section. First, all consumers are homogeneous and equally likely to enjoy a good of the same quality. The product quality $\theta \in \{h, \ell\}$ can be either high (h) or low (ℓ). A high-quality product has a probability $p_h \in [0, 1]$ of being enjoyed by a consumer, while a low-quality product has a probability $p_\ell \in [0, 1]$ of being enjoyed, with $p_h > p_\ell$.

The set of products purchasing fake reviews is also assumed to be characteristically identical to those that do not. Thus, $p_\ell = .725$ and $p_h = .875$, and the prior distributions $p(p_h)$ and $p(p_\ell)$ are Beta$(1,1)$ since it is assumed that there is an equal proportion of high and low-quality products in the population.

Similar to the previous model, an exogenous number of sales q are made, and following q sales, exactly αq reviews are posted for $\alpha \in (0,1]$. Each review indicates whether or not the product was enjoyed by the consumer and provides no further information. Additionally, X_h and X_ℓ are random variables that record whether a reviewer enjoyed a product with quality h and ℓ, respectively, and do not include any fake reviews. Reviews are sampled in the same manner as in Section 4, so $p(X_h|p_h) \sim \text{Bin}(\alpha q, p_h)$ and $p(X_\ell|p_\ell) \sim \text{Bin}(\alpha q, p_\ell)$.

Finally, firms have the option to purchase fake reviews before making any sales. These reviews are indistinguishable from truthful reviews and are always positive. The number of fake reviews purchased by a high and low-quality firm is denoted as s_h and s_ℓ, respectively, and both values are non-negative. The cost of purchasing s fake reviews is $C(s)$, where $C'(s) > 0$ and $C''(s) > 0$. This cost becomes sufficiently large such that firms are not incentivized to purchase an infinite number of fake reviews.

As shown in the appendix, $p(p_h|X_h) \sim \text{Beta}(X_h + 1 + s_h, \alpha q - X_h + 1)$ and $p(p_h|X_\ell) \sim \text{Beta}(X_\ell + 1 + s_\ell, \alpha q - X_\ell + 1)$.

5.2 Model with High and Low-Quality Fakers

I investigate the number of reviews (both fake and organic) required to reveal product quality as the number of fake reviews purchased by high and low-quality firms varies.

5.2.1 Monte Carlo Simulations

Using the methodology detailed in Section 4, I conduct Monte Carlo simulations following the choice of four different amounts of review fraud for each product type: $0, 3, 8,$ and 20 fake reviews. While firms can choose to purchase any amount of fake reviews, these values represent a firm that purchases a small, medium, and large number, as well as none.

As He et al. (2022) calculate, products with low markups require about ten organic sales to offset the cost of one fake review, while those with high markups need only one sale. So, I surmise that new products seeking fake reviews would purchase at least a handful of reviews, but would avoid purchasing so many they are caught by the FTC or Amazon.

Since both high and low-quality products can choose any of these four levels, there are 16 possible environments to consider. For each combination, I generate the distribution of the number of reviews needed to reveal product quality. Figure 6 and Table A.2 provide a histogram and descriptive statistics for all possible combinations.

As expected, the number of reviews required to reveal quality increases as the number of fake reviews purchased by the low-quality firm rises, and decreases as s_h rises. These results are intuitive: a higher number of fake reviews purchased by a low-quality good makes it harder to distinguish between different product qualities, while more review fraud for a high-quality firm allows consumers to correctly "jump to a conclusion." Moreover, as both types of firms purchase more fake reviews, the number of total reviews needed to reveal product quality increases.

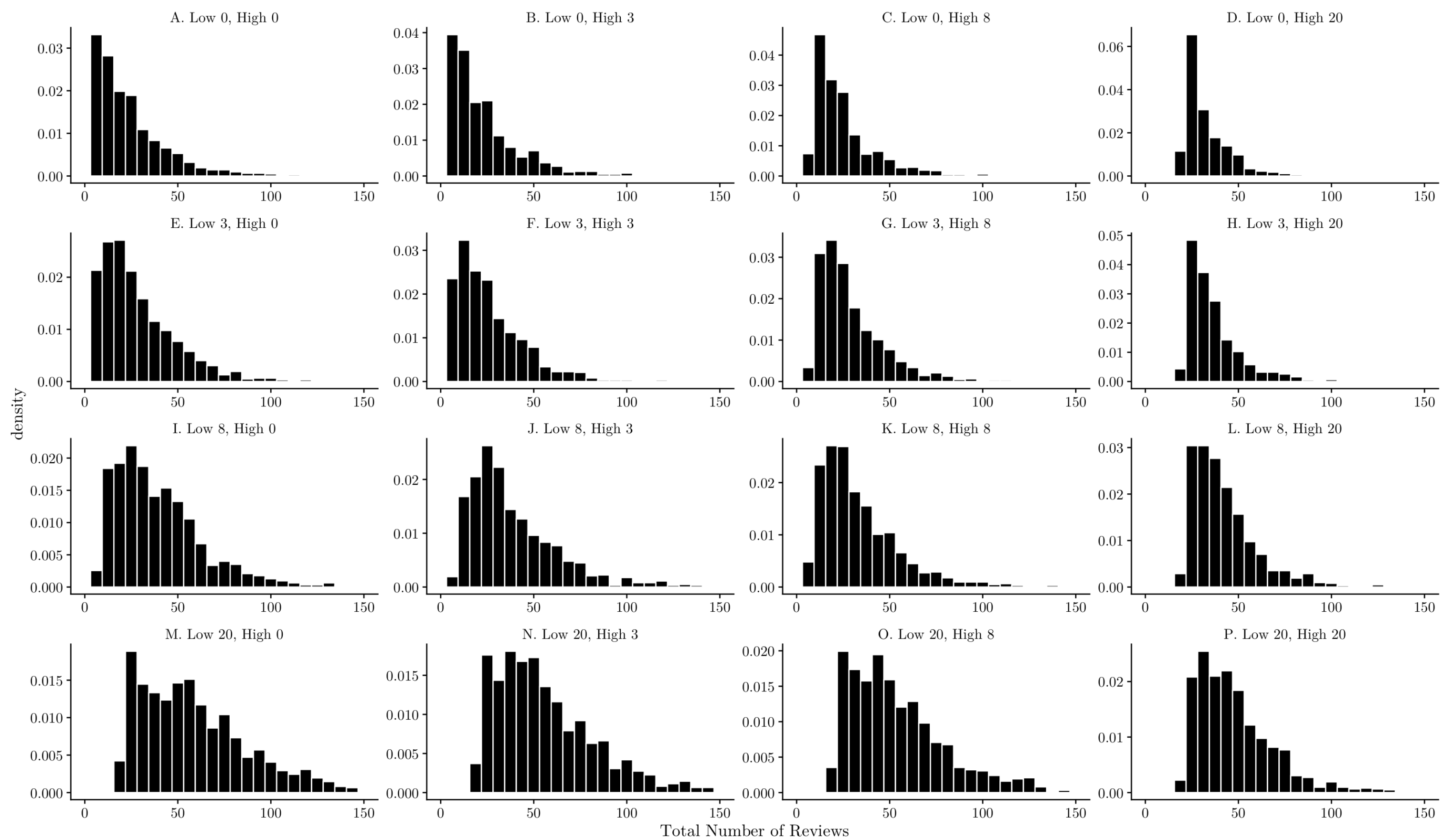
5.3 Empirical Estimation of Review Fraud

In this section, I use similar empirical methods as outlined in Section 4 to determine the number of reviews necessary to reveal product quality and attempt to estimate s_h and s_ℓ.

5.3.1 Reviews to Reveal Quality

I consider a new product on Amazon with little public information. To estimate the minimum number of reviews required to reveal quality, I must determine how many reviews are typically needed for the average star rating to converge. This occurs when the change in average rating is very small following a new review. I use .025 as the necessary threshold to determine if a new review does not significantly affect the average rating, and Figure A.4 shows the average absolute change in average rating per review as more reviews are posted.

Figure 6: Simulated Distributions for Firms Purchasing Fake Reviews



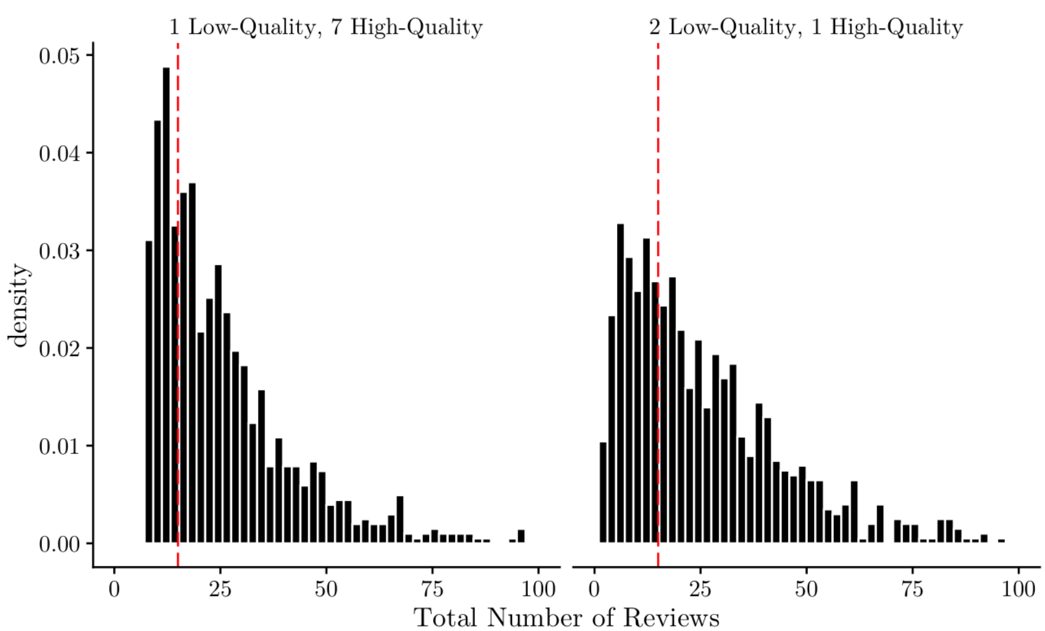
Using a one-sample Wilcoxon signed rank test, I determine the minimum number of reviews such that the median change is statistically significantly less than .025. When 21 reviews are posted, the null hypothesis is rejected at the 5% significance level, implying that the number of reviews required to distinguish high and low-quality products is 21. This is slightly more than the quality-revealing number of reviews among the control products (20).

According to these empirical results, more reviews are needed for consumers to detect product quality in the presence of fake reviews, a result that corroborates some previous literature and reveals potential welfare losses due to the existence of fake reviews.

5.3.2 Estimating Fake Reviews

Given that 21 reviews are necessary to differentiate between high and low-quality products, I present two different estimations for s_ℓ and s_h. First, $s_\ell = 2$ and $s_h = 1$. Alternatively, $s_\ell = 1$ and $s_h = 7$. For both estimates, I conduct 1000 repeated Monte Carlo simulations and find that for each simulated distribution, the median number of reviews is 21 (Table A.3). Figure 7 depicts a histogram for each simulated distribution.

Figure 7: Simulated Review Distributions for Estimates of Fake Reviews



Based on this model, the empirical results, and the two possible predictions for s_ℓ and s_h, I cannot definitively conclude whether high-quality firms purchase more or less fraudulent reviews than their low-quality counterparts.

6 Empirical Framwework and Results

In this section, I investigate the inconclusive results produced by the model in Section 5. In the absence of review fraud, it takes slightly fewer reviews for a product to converge to its true quality, but the number of fake reviews purchased by a high and low-quality firm cannot be determined. To test these findings, I formulate two empirical models.

6.1 Identifying High and Low-Quality Products

In Section 3, I identify 796 products in the focal group as being new to the marketplace. Among these products, I select the ones with greater than 10 total reviews and examine the (at most) 20 most recent reviews for each good that are posted at least two weeks after it stops seeking fake reviews on Facebook. For most products, these reviews are written many weeks following the first instance of review fraud, so I propose that they are unbiased indicators of the product's true quality.

Using p_h and p_l as calculated in Section 4, I classify product quality as follows: if the 20 most recent reviews have an average above 4.5 stars, the good is high-quality. If the average of the most recent reviews is below 3.9 and above 2.5, then the good is low-quality. Products that have an average outside of these cutoffs are identified as "unknown." Using this methodology, 249 products are identified as high-quality, 142 are low-quality, and for 238 goods, the quality is not able to be determined.

6.2 Empirical Framework

Given that organic and fraudulent reviews appear nearly identical, it is difficult to identify and count the number of fake reviews purchased by each good. However, firms purchase these reviews through Facebook groups and I observe the first and last Facebook post for each product. Therefore, I use the number of full weeks in which a product seeks fake reviews (*postinglength*) as a proxy for the number of fake reviews purchased and conjecture that the more weeks in which a product posts on Facebook, the more fake reviews it buys. Posting on Facebook for a longer period of time is not conclusive evidence that the firm purchased more fake reviews. There are many other reasons why this might occur that are not captured in the data such as a language barrier between the firm and a reviewer.

Several product features that are observed in the data can also affect the posting length on Facebook and are thus controlled for in the following empirical estimation models. The number of reviews (*reviews*), price (*price*), and the position at which the product appears when searched for on Amazon when beginning to seek fake reviews (*position*) would all affect the posting length. I propose the following Ordinary Least Squares (OLS) regression models to estimate the effect of quality on the number of fake reviews purchased by product i and apply a log transformation on *postinglength* since it is right-skewed.

$$\ln(postinglength_i) = \beta_1 newavg_i + \beta_2 reviews_i + \beta_3 price_i + \beta_4 position_i \tag{1}$$

$$\ln(postinglength_i) = \gamma_1 high_i + \gamma_2 reviews_i + \gamma_3 price_i + \gamma_4 position_i \tag{2}$$

The primary variables of interest are *newavg* and *high*. The first specification examines the effect that the recent average (*newavg*), without classification into high and low-quality, has on the posting length on Facebook. The second classifies products into high and low-quality (*high*), as explained above.

Additionally in (2), all observations in which quality cannot be determined are removed. In both models, I remove outlier observations where the length of Facebook posting is greater than 14 weeks, as calculated from Table 3. Observations in which *position* and *price* were not calculated are also removed. Regression results are provided in Table 4.

6.3 Regression Model Results

In both specifications, unlike the inconclusive results produced by the model, the presumed quality is a statistically significant predictor of the length of Facebook posting. When the quality of the good was lowered, all else constant, the longer the product posted on Facebook. Specifically, I estimate that an increase in the true product quality by 1 star decreases the length of review fraud by about 13% and a high-quality product will post on Facebook for roughly 28% less time than a low-quality counterpart, ceteris paribus. Besides estimated quality, *position* and *price* were also statistically significant predictors in the first specification; however, they both had a small impact on the posting length. These results validate the underlying assumption of this paper that a firm's decision to purchase fake reviews depends on its quality.

Table 4: Estimation Results of Regression Models

	(1)	(2)
	$\ln(postinglength)$	$\ln(postinglength)$
newavg	-0.1325**	
	(0.0546)	
high		-0.2811***
		(0.0940)
reviews	0.0001	0.0001
	(0.0001)	(0.0001)
price	-0.0022**	-0.0010
	(0.0010)	(0.0013)
position	0.0011**	0.0008
	(0.0005)	(0.0006)
Observations	447	277
R^2	0.041	0.046

Standard errors in parentheses

Data source: He et al. (2022)

$^{*}\ p < .10, ^{**}\ p < .05, ^{***}\ p < .01$

7 Conclusion

In this paper, I propose a theoretical framework that simulates the minimum number of reviews required to differentiate between high-quality and low-quality products on eCommerce sites like Amazon in the presence of review fraud. To estimate various parameters of the model, including the number of fake reviews purchased by each product type, I use data from He et al. (2022) that identifies and tracks products on Amazon that buy fake reviews through Facebook. I find that the focal group of products experience a greater change in their average rating compared with the control group and conclude that it takes one more review for quality to be determined among products that commit review fraud. However, the results from the model are inconclusive and cannot determine whether high-quality products buy more or fewer fake reviews than their low-quality counterparts.

Furthermore, I empirically test the results of the model using two OLS regression specifications and examine the relationship between perceived quality and the estimated amount of review fraud. I predict a firm's quality by analyzing its most recent reviews and approximate the amount of review fraud by calculating the number of weeks in which a firm seeks fake reviews on Facebook. Unlike the predictions from the model, these empirical results show that high-quality firms tend to purchase fewer fake reviews than low-quality ones, holding all other variables constant.

The model has many limitations that could be improved upon in future research. First, the assumption that there are only two product qualities is a simplification that does not accurately reflect the complexity of the product market on Amazon. The proposition that the population is the same among products that do and do not purchase fake reviews is also problematic, as this treatment is not randomly assigned. Additionally, the model assumes that the proportion of high and low-quality products on Amazon is the same. However, as shown in Figure 5, the majority of products have an average star rating above 3.5. In future models, if this proportion is variable, then products with extremely low ratings can be included.

Finally, the model assumes that sales are exogenous and identical for both products. In future work, sales could be a function of social learning, where each consumer decides independently to buy the product based on its reviews. Following a sale, this consumer posts a new review containing information that is used by all agents in the model. This process could continue for a set amount of time until the products are revealed to be different in quality.

Future research can also expand on the empirical methodology. The empirical analysis conducted in Section 5 is not causal. There are many other reasons why convergence in average rating occurs quicker in the control group, such as the population quality being different than the focal group or an error in data collection. Additionally, the standard deviations of the simulated distributions are quite large, raising the possibility that the observed "trial" in the data may not be representative of the population distribution. Finally, in Section 6, I estimate product quality and the amount of review manipulation, but cannot directly compute these. As algorithms and artificial intelligence become better at detecting fake reviews, future work can more directly establish the relationship between quality and review fraud.

Despite the limitations of this paper, I believe that this type of analysis is important to continue. Several theoretical models of review fraud have been created, but very few have been empirically tested. As data on review fraud becomes increasingly accessible, proposed models must be analyzed for their applicability to the current eCommerce landscape.

8 Appendix

8.1 Additional Empirical Work

Figure A.1: McCrary Test for Average Star Rating at 4.25 (All Products)

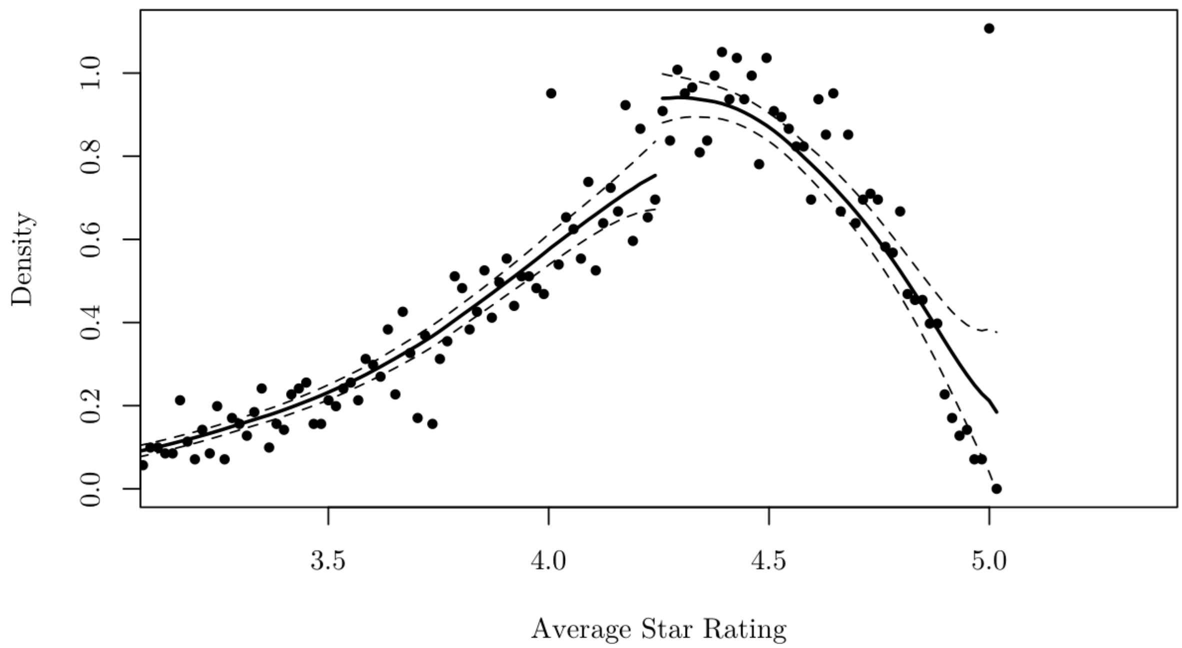


Figure A.2: Mean and Median Reviews to Reveal Quality

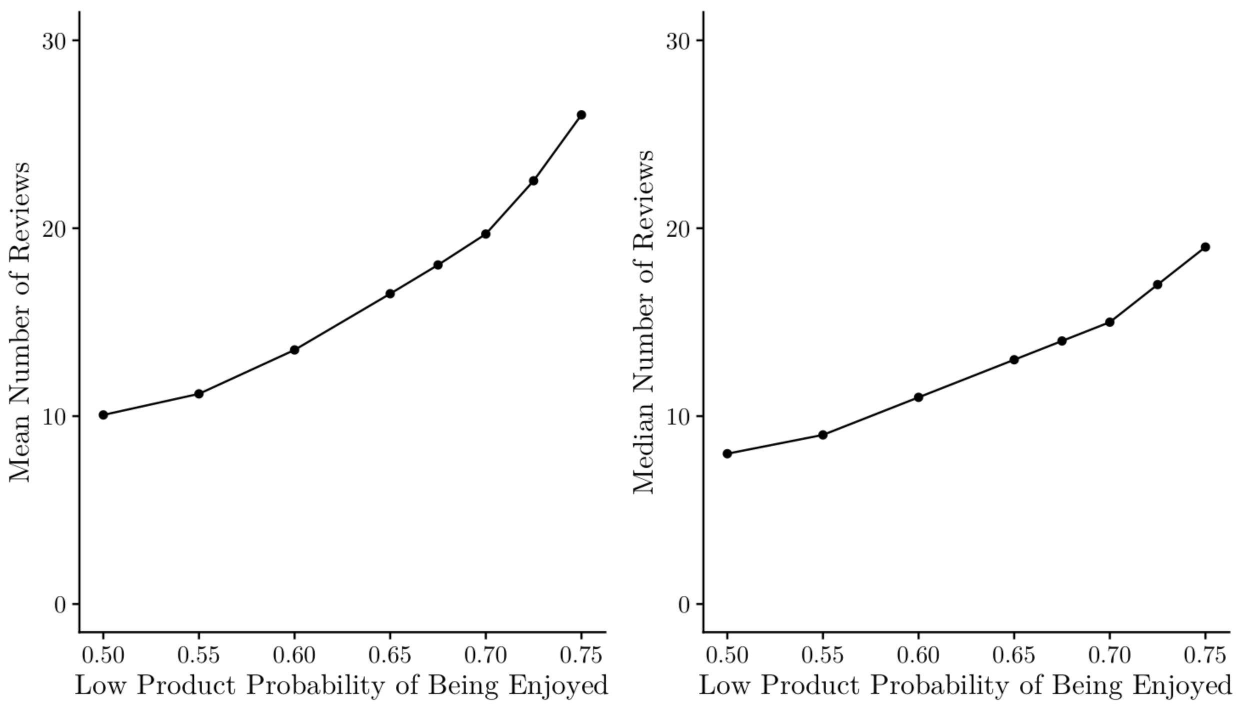


Figure A.3: Average Change in Average Rating (Control Products)

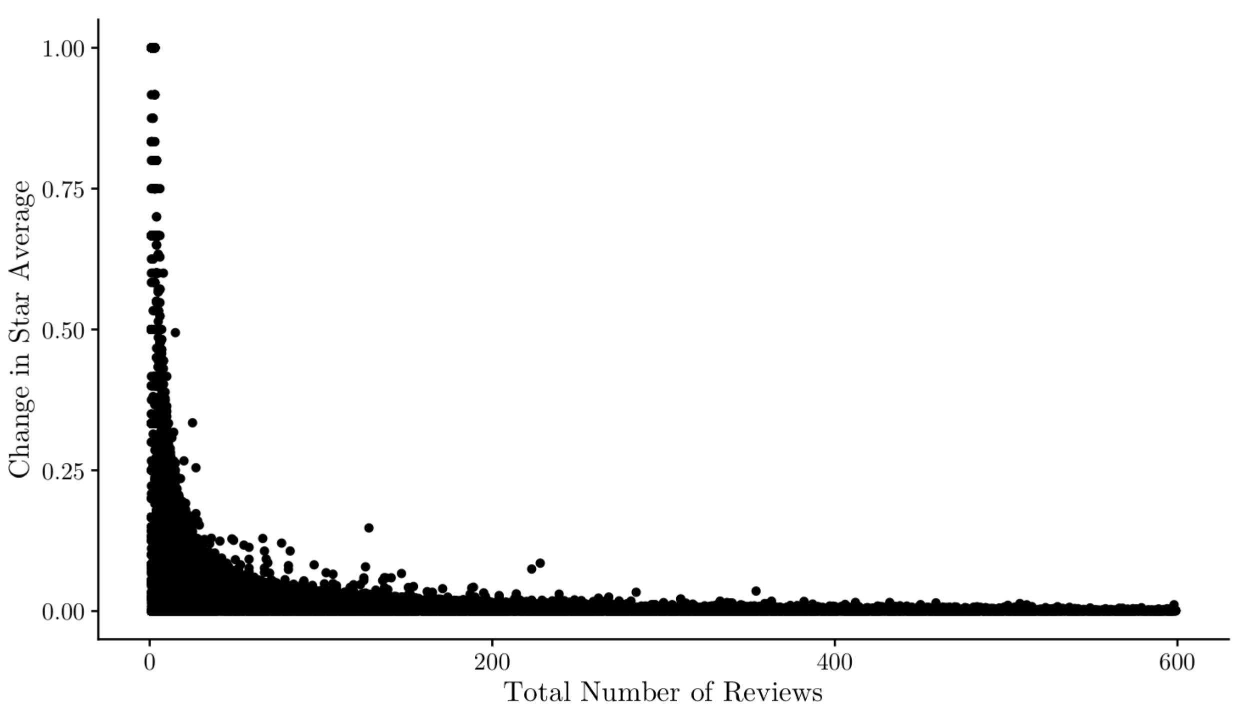


Table A.1: Summary Statistics for Simulated Review Distribution (No Fake Reviews)

p_h and p_l	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
$p_h = .875$ and $p_l = .725$	24.11	23.01	8.00	18.00	32.00

Table A.2: Summary Statistics for Simulated Review Distributions

Fake Reviews	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Low 0, High 0	21.93	20.44	7.00	16.00	31.00
Low 0, High 3	23.47	20.28	10.00	17.00	30.00
Low 0, High 8	25.62	17.79	13.00	20.00	31.00
Low 0, High 20	33.27	13.29	24.00	29.00	38.00
Low 3, High 0	28.79	21.08	14.00	23.00	38.00
Low 3, High 3	26.23	18.67	13.00	21.00	35.00
Low 3, High 8	29.35	18.52	16.00	24.00	37.00
Low 3, High 20	36.64	14.65	27.00	33.00	42.00
Low 8, High 0	39.34	23.33	21.00	34.00	51.00
Low 8, High 3	38.81	24.03	22.00	33.00	50.00
Low 8, High 8	33.65	21.88	18.00	28.00	43.00
Low 8, High 20	42.14	17.17	30.00	38.00	49.25
Low 20, High 0	60.83	32.35	36.00	55.00	77.00
Low 20, High 3	56.65	27.83	36.00	50.50	72.00
Low 20, High 8	54.85	28.27	34.00	48.50	67.00
Low 20, High 20	47.93	20.84	32.00	43.00	59.00

Figure A.4: Average Change in Average Rating (Faking Products)

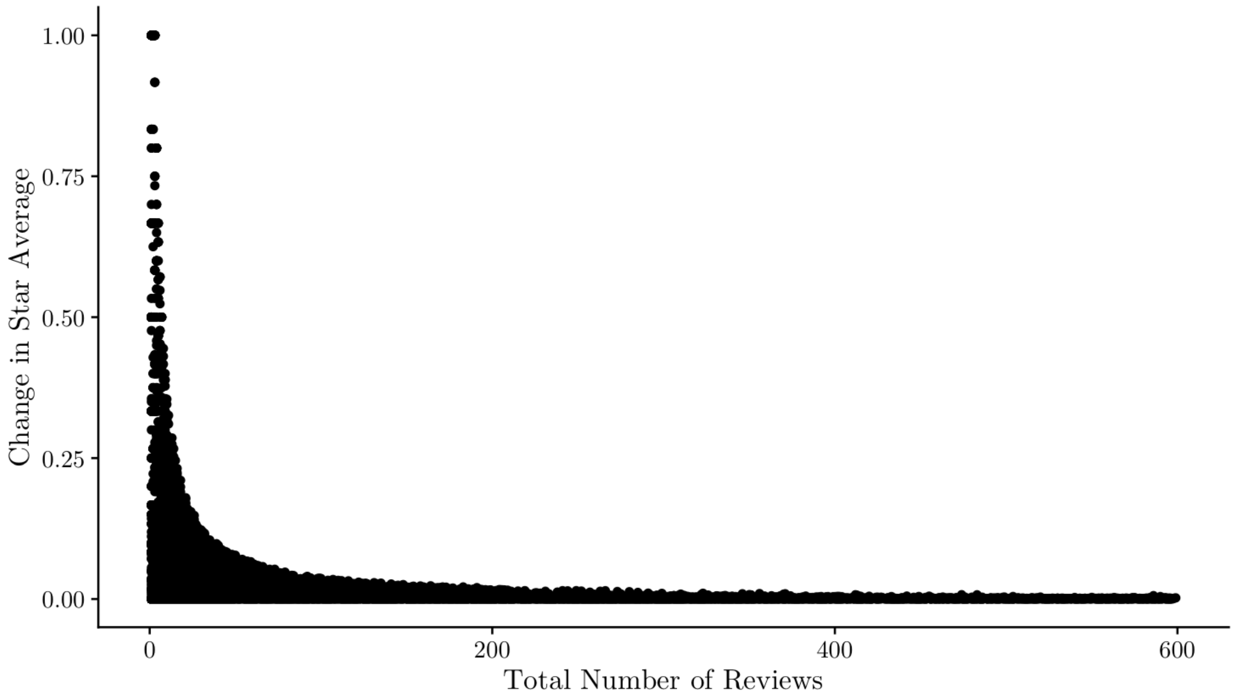


Table A.3: Summary Statistics for Simulated Review Distributions (with Fake Reviews)

s_l and s_h	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
$s_l = 1$ and $s_h = 7$	26.00	17.74	13.75	21.00	32.00
$s_l = 2$ and $s_h = 1$	27.06	22.11	12.00	21.00	36.00

8.2 Proof of Posterior Distribution

Let X be a random variable with $X \sim \mathrm{Bin}(n, p)$. Suppose $\pi(p)$ is a random variable with $\pi(p) \sim \mathrm{Beta}(1, 1)$. This is will be a proof of the posterior distribution $p(p|X)$. By Bayes Law:

$$p(p|X) = \frac{f(X|p) \star \pi(p)}{\int f(X|p) \star \pi(p) dp} \tag{1}$$

Since $\pi(p) \sim \mathrm{Beta}(1, 1)$ and $X \sim \mathrm{Bin}(n, p)$, it is known that $\pi(p) = 1$ and $f(X) = p^X(1-p)^{n-X}$. So, (1) can be simplified:

$$p(p|X) = \frac{p^X(1-p)^{n-X}}{\int_0^1 p^X(1-p)^{n-X} dp} \tag{2}$$

Since the denominator of (2) is a constant term, $p(p|X) \propto p^X(1-p)^{n-X}$. Thus, $p(p|X) \sim \mathrm{Beta}(X + 1, n - X + 1)$.

References

Anderson, S., & Renault, R. (2006). Advertising Content. *American Economic Review*, *96*(1), 93–113.

Bagwell, K. (2007). Chapter 28 The Economic Analysis of Advertising. In M. Armstrong & R. Porter (Eds.). Elsevier.

Banerjee, A., & Fudenberg, D. (2004). Word-of-mouth learning. *Games and Economic Behavior*, *46*(1), 1–22.

FTC to Explore Rulemaking to Combat Fake Reviews and Other Deceptive Endorsements. (2022). https://www.ftc.gov/news-events/news/press-releases/2022/10/ftc-explore-rulemaking-combat-fake-reviews-other-deceptive-endorsements

Glazer, J., Herrera, H., & Perry, M. (2018). *Fake Persuasion* (CEPR Discussion Papers No. 13244). C.E.P.R. Discussion Papers.

Hattori, K., & Higashida, K. (2012). Misleading advertising in duopoly. *Canadian Journal of Economics/Revue canadienne d'économique*, *45*(3), 1154–1187.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The Market for Fake Reviews. *Marketing Science*, *41*(5), 896–921.

Hollenbeck, B., Moorthy, S., & Proserpio, D. (2019). Advertising Strategy in the Presence of Reviews: An Empirical Analysis. *Marketing Science*, *38*(5), 793–811.

Kelly, D., & Kolstad, C. (1999). Bayesian learning, growth, and pollution. *Journal of Economic Dynamics and Control*, *23*(4), 491–518.

Kihlstrom, R. E., & Riordan, M. (1984). Advertising as a Signal. *Journal of Political Economy*, *92*(3), 427–50.

Lee, D. S., & Moretti, E. (2009). Bayesian Learning and the Pricing of New Information: Evidence from Prediction Markets. *American Economic Review*, *99*(2), 330–36.

Luca, M. (2011). Reviews, Reputation, and Revenue: The Case of Yelp.Com. *SSRN Electronic Journal*.

Luca, M., & Zervas, G. (2016). Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science*, *62*(12), 3412–3427.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, *104*(8), 2421–55.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714.

Milgrom, P., & Roberts, J. (1986). Price and Advertising Signals of Product Quality. *Journal of Political Economy*, *94*(4), 796–821.

Piccolo, S., Tedeschi, P., & Ursino, G. (2015). How limiting deceptive practices harms consumers. *The RAND Journal of Economics*, *46*(3), 611–624.

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

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, *9*(2), 185–202.

Smith, A., & Anderson, M. (2016). *2. Online reviews*. https://www.pewresearch.org/internet/2016/12/19/online-reviews/

Sobel, J. (2000). Economists' Models of Learning. *Journal of Economic Theory*, *94*(2), 241–261.