Misinformation and Mistrust: The Equilibrium Effects of Fake Reviews on Amazon.com

Ashvin Gandhi
Anderson School of Management
UCLA *

Brett Hollenbeck
Anderson School of Management
UCLA †

February, 2024 ‡

This paper investigates the impact on consumers of the widespread manipulation of reputation systems by sellers on two-sided online platforms. We focus on a relevant empirical setting, the use of fake product reviews on e-commerce platforms, which can affect consumer welfare via two channels. First, rating manipulation deceives consumers directly, causing them to buy lower quality products and pay higher prices for the products with manipulated ratings. Second, the presence of rating manipulation lowers trust in ratings, which may result in worse product matches if consumers place too little weight on quality ratings. This decrease in trust may also increase price competition and benefit consumers by lowering prices on high quality products whose quality is less easily observed. We formally model how consumers form beliefs about quality from product ratings and how these beliefs are affected by the presence of fake reviews. We use incentivized survey experiments to measure beliefs about fake review prevalence. Our model of product quality is incorporated into an empirical model of consumer demand for products and how demand is shifted by ratings, reviews, and prices. The model is estimated using a large and novel dataset of products observed buying fake reviews to manipulate their Amazon ratings. We use counterfactual policy simulations in which fake reviews are removed and consumer beliefs adjust accordingly to explore the effectiveness and welfare and profit implications of different methods of regulating fake reviews.

* ashvin.gandhi@anderson.ucla.edu

† brett.hollenbeck@anderson.ucla.edu

‡ We thank the Morrison Center for Marketing Analytics for generous funding and Zhijian Li for excellent research assistance. We thank Joel Waldfogel and Jessica Fong for helpful comments, as well as seminar participants at UPenn, UT Austin McCombs, Santa Clara University, UCSD Rady, UC Riverside, the University of Toronto - Rotman, Yale SOM, the 2024 Bass Forms Conference, and the 2023 FTC Microeconomics Conference.

1 Overview

The use of user generated ratings and reviews as a tool for sellers to establish reputations has proven to be essential for the success of two-sided marketplaces such as e-commerce platforms (Cabral and Hortacsu (2010), Tadelis (2016), Einav et al. (2016)). Survey data show that an overwhelming majority of consumers consult reviews before making purchases, both online and offline. The information provided by reviews and their ability to improve match quality not only benefit sellers and platforms, but consumers as well (Reimers and Waldfogel (2021), Wu et al. (2015)).

An implication of this is that, by helping consumers learn product quality, a product's ratings necessarily have a significant impact on their sales. Consequently, sellers have a strong incentive to manipulate these ratings and recent research has documented that rating manipulation using fake, paid-for reviews is quite common and arguably more so now than ever ((He et al., 2022b), (Commission, 2023)). The use of fake reviews by sellers has become an important issue for regulators. The FTC, the UK CMA, the European Commission, and others have all investigated into the potential consumer harms from rating manipulation and proposed laws or other measures in response.[1] In this paper, we therefore study how consumers, sellers, and platforms are impacted by rating manipulation.

We start by laying out a formal model of the consumer welfare effects of rating manipulation. Fake reviews impact consumer welfare via two primary channels. The first channel (which we refer to as "misinformation") is that fake reviews mislead consumers by inflating the ratings of certain products, leading consumers to make sub-optimal purchases. They may cause consumers to purchase products that are of lower quality than those they would have purchased in the absence of fake reviews, or else to pay a higher price due to the inflated ratings. The second channel (which we refer to as "mistrust") is that awareness that some reviews are fake will lead rational consumers to discount the credibility of good reviews. This

[1] See:https://www.ftc.gov/news-events/press-releases/2019/02/ftc-brings-first-case-challenging-fake-pa and https://www.gov.uk/government/news/cma-investigates-misleading-online-reviews.

will also serve to shift demand from high-quality to low-quality products. One consequence of less trust in ratings, however, is that with a lower consumer elasticity to ratings high-quality products are less able to differentiate themselves from low-quality products through higher ratings and therefore must compete more aggressively on price. Notably, if most consumers still buy the high-quality product, the result of this increase in price competition is to increase their welfare.The relative magnitude of these different forces and thus the net impact of fake reviews on aggregate welfare is therefore ambiguous.[2]

This possibility for the presence of misinformation to increase welfare has been shown theoretically in the analogous setting of false advertising (Rhodes and Wilson (2018)), who show conditions where there exist equilibria with positive amounts of false advertising that can increase consumer welfare by increasing price competition between high- and low-quality products. Other theoretical work on rating manipulation suggest it could be welfare improving (Dellarocas (2006), Glazer et al. (2020), Saraiva (2020), and Yasui (2020)).[3]

We bring novel data to the debate over the nature and magnitude of welfare harm from fake reviews. We follow on earlier research (He et al. (2022b)) that documents the widespread purchasing of fake reviews by product sellers on Amazon in private Facebook groups.[4] We use novel data on this market, in which a set of research assistants joined and observed private Facebook groups where sellers solicit fake reviews, and constructed a set of roughly 1,500 products known to use fake reviews, as well as the timing of their fake review campaigns. For these products and a set of competitors, a large-scale panel of data was also collected from Amazon on their ratings, reviews, sales ranks, prices, and advertising behavior. We

[2] A third channel by which seller manipulation of ratings may impact consumers is through dynamic effects, namely the extent to which paying for reviews lowers barriers to entry for high-quality entrants, or alternatively the extent to which low trust in reviews increases the barriers to entry for this type of seller. These types of dynamic effects are outside the scope of this paper.

[3] In a similar setting as ours Li et al. (2020) shows that allowing sellers to buy reviews can increase market efficiency because higher quality sellers are more likely to buy reviews than lower quality sellers. In this case, however, the necessary separating equilibrium requires that sellers commit to paying buyers for reviews regardless of their sentiment (i.e. reviewers can leave negative reviews for low-quality products). In our setting, we document that the fake reviews are required to be positive for the reviewer to be compensated.

[4] While we focus on fake Amazon reviews, similar marketplaces exist for other e-commerce platforms like Wayfair, Walmart, Yelp, and so on.

combine these pieces and other data sources to measure the impact of ratings and reviews on consumer demand.

We first model how consumers form beliefs about product quality from ratings, where these beliefs incorporate the existence of fake reviews. Organic reviews are treated as reflecting product quality whereas fake reviews are positive by assumption. Consumers form their expectations from a Bayesian model of product quality given the observed reviews. We show that consumers' knowledge that fake reviews exist cause them to trust reviews less, believing that products have lower quality than if they had the same rating in a world without fake reviews.

Estimating this model empirically requires us to take a stance on what consumers believe about how common fake reviews are. To inform our assumed values on these beliefs, we conduct a set of large-scale incentive-compatible survey experiments designed to elicit beliefs about fake reviews in the population of Amazon shoppers, as well as how accurately they can detect which products use fake reviews. We find that survey respondents have roughly accurate beliefs about how prevalent fake reviews are, but do badly at identifying which products use them. Finally, we show in an appendix how our results vary for alternative assumptions on beliefs such as rational expectations.

Next, we estimate a structural model of demand following Berry et al. (1995), taking as given that consumers' beliefs about product quality from ratings factor into demand. This models consumer demand is a function of their beliefs over product quality derived from ratings and not simply the ratings themselves. An implication is that the same rating can yield different demands depending on consumers' beliefs about the presence of fake reviews. Our demand estimation produces reasonable values of price elasticities and of the elasticity of demand with respect to perceived product quality. But because our estimates come from a structural model of demand and beliefs, we can construct counterfactual demand under alternative ratings regimes.

To measure the impact of fake reviews on consumer welfare, we consider a series of counterfactual policy analyses that isolate the different mechanisms at play. We use our knowledge of which products use fake reviews, as well as estimates of the proportion of their reviews that are fake, to adjust downward their average ratings and number of reviews as if the platform had deleted all fake reviews. We then recompute equilibrium prices and calculate consumer welfare and firm profits when fake reviews are present vs when they are absent. In addition, we isolate the misinformation and mistrust effects by simulating demand under partial counterfactuals. In the first, we isolate misinformation by reintroducing fake reviews and setting consumers' beliefs to be fully trusting in reviews. Next, we isolate mistrust by removing fake reviews but setting consumers' beliefs as if they were still present. In both cases we show results with fixed prices and with competitive reactions in order to understand the role of price adjustments in each outcome.

We find evidence for substantial consumer harm from sellers manipulating their ratings. The effects are on the order of a roughly 2.5% across-the-board increase in prices for all products. Price competition and consumer beliefs about the trustworthiness of ratings plays a large role in this result. When isolated, the misinformation effect of fake reviews is to cause a modest decrease in consumer welfare as consumers are led to buy lower quality products. By contrast, when fake reviews are not present but consumers are mistrustful of ratings, and thus relatively insensitive to ratings, consumers make worse purchases and are substantially worse off. When both fake reviews are present and consumers are aware of this, honest sellers react to the increase in competition by decreasing prices and this partially offsets the large welfare harms from misinformation and results in a fairly modest overall decrease.

Finally, we find that the platform would benefit significantly in the full counterfactual if it deleted all fake reviews and shifted consumer beliefs accordingly. If it simply deletes fake reviews without consumers adjusting their beliefs, however, platform revenue falls. From the platform's perspective, that is, simply removing fake reviews could backfire in the short run if consumers are not informed about this or do not find it credible.

We contribute to several strands of literature related to information disclosure, platform design, and reputation manipulation. First, and most directly, we contribute to the growing literature on fake reviews which begins with Mayzlin et al. (2014) and Luca and Zervas (2016). Theoretical work on fake reviews has shown that under reasonable circumstances, fake reviews can be efficient and welfare-enhancing. In an extension of the signal-jamming literature on how firms can manipulate strategic variables to distort beliefs, Dellarocas (2006) 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 (see Glazer et al. (2020), Saraiva (2020), and Yasui (2020).) These are full equilibrium models of the seller decision to use fake reviews in which consumer beliefs rationally forecast equilibrium seller behavior. Our theoretical framework instead allows consumers to have a range of beliefs, including being naive with respect to the presence and prevalence of fake reviews, but as a consequence should be thought of as a partial equilibrium model.

There have been few attempts to empirically test or quantify the predictions of these models or to empirically assess the impact of fake reviews on welfare or competition. An exception is Akesson et al. (2022), who conduct an incentive-compatible online experiment in which products are shown with random variation in whether and how fake reviews appear. They find via choice tasks that the presence of fake reviews makes consumers more likely to purchase lower-quality products and estimate a welfare loss of $.12 for each dollar spent from this mechanism. This experiment therefore captures the direct effect of misinformation, but does not try to quantify the indirect effects of the change in equilibrium prices that result and does not address the effects of mistrust. Another closely related work is Li et al. (2020), an examination of incentivized reviews on Taobao. They find that high-quality sellers select into the incentivized review system and this improves market efficiency. There are several distinguishing features of incentivised reviews, compared to fake reviews, that we describe in more detail below. While not considering fake reviews, Reimers and Waldfogel (2021) study the welfare impact of consumer reviews as a whole, showing that Amazon user reviews have a large impact on consumer surplus.

We also contribute to an emerging literature on information disclosure. Dranove and Jin (2010) summarize a large body of research on quality disclosure, with a focus on voluntary firm disclosure. When a platform acts as an intermediary and designs a system of quality disclosure, new and complex incentives around competition and welfare arise.[5] Armstrong and Zhou (2022) provide a general treatment of the issues around information signals and competition, and show that a policy that dampens differentiation can intensify competition and benefit consumers.[6] Hopenhayn and Saeedi (2023) characterize an optimal rating system in the presence of competition and adverse selection by sellers. They show that more precise quality ratings does not always benefit consumers. In ongoing work, Saeedi and Shourideh (2020) studies optimal ratings when firms can potentially manipulate ratings. Vatter (2021) also shows that full information disclosure is not optimal, and characterizes optimal quality scores in the context of Medicare Advantage. Our contributions to this literature are, first, to show how endogenous mistrust of disclosed information could produce similar results as coarse disclosure, and second, empirically characterizing whether consumers are better off by placing less trust in quality ratings.

2 Data

This section describes the data used in our analysis and summarizes some of its key features. To measure the effects of rating manipulation on consumer demand and understand how that demand would change under alternative regulatory scenarios, we require data from Amazon on product characteristics, ratings, sales, and how these vary over time, as well as

[5] Notable related work on platform reputation systems includes Dai et al. (2018), Hui et al. (2016), Hui et al. (2022), and Chakraborty et al. (2022).

[6] Related work by Vellodi (2018) focuses on dynamics, and shows that suppressing the reviews of highly-rated firms can stimulate entry and improve consumer welfare through that channel.

information on which products are using fake reviews and their extent of fake review activity.

The primary channel where sellers obtain fake reviews is a set of private Facebook groups (He et al., 2022b), which operate in the following way. Sellers post a photo of their product and solicit reviews from interested reviewers, who then engage in a private conversation with the seller. The reviewer then purchases the product and leaves an authentic-seeming "verified purchase" review, after which they are compensated via a PayPal payment in the amount of the purchase price plus any taxes and fees and, in some cases, a small commission. This compensation is contingent on the review being positive with a five-star rating and evading any filtering algorithms used by Amazon to prevent fake reviews.

Note that the practice of purchasing fake reviews differs from the sanctioned use of "incentivized reviews." As with fake reviews, sanctioned incentivized reviewers do receive the product for free or at a discount. However, unlike fake reviews, incentivized reviews must clearly disclose this arrangement, and incentivized reviewers receive the same payment for positive and negative reviews. Moreover, Amazon's incentivized review program (known as Amazon Vine) does not allow sellers to choose their own incentivized reviewers.

We obtain data on fake review activity by collecting information directly from the private Facebook groups where fake reviews are bought by product sellers. As scraping Facebook is technically infeasible, this required using a team of research assistants to hand-collect data on what products were posting in the Facebook groups and during what times they were actively recruiting fake reviews there. More information on these groups and how the data were collected are described in detail in He et al. (2022b). Our data collection provides us with information on a set of roughly 1,500 unique products observed buying fake reviews between October 2019 and June 2020.

In addition, we conduct a large-scale scraping of Amazon.com repeatedly during and after this time period. This scraping is centered around searches of the product keyword as identified by the seller. For each keyword, on each day we collect the full set of product results including the products' positions in the search results, their prices, number of reviews, average rating, and presence of sponsored links. We use the keyword results to define for each product a set of close competitors. These are defined as the products that show up most frequently near the focal product in the search results around the time the focal product begins soliciting fake reviews. For both the focal products and this set of close competitors, we repeatedly collect the complete history of their reviews including the text and photos used in each review. For every product review, we also collect the reviewer ID and use this to compile the set of other products also reviewed by these reviewers. This will be useful later in estimating the share of fake reviews for each focal product.

Product Information Table 1 reports the top product categories and subcategories in the dataset. Notably, products using fake reviews are found across a wide range of categories and subcategories.

Table 1: **Top Categories and Subcategories**

Category	N	Subcategory in top category	N
Beauty & Personal Care	598	Men's Rotary Shavers	37
Health & Household	525	Hot-Air Hair Brushes	34
Tools & Home Improvement	387	Light Hair Removal Devices	33
Home & Kitchen	374	Blemish & Blackhead Removal Tools	27
Kitchen & Dining	352	Facial Masks	26
Cell Phones & Accessories	236	Makeup Brush Cleaners	22
Pet Supplies	211	Eyelash Curlers	21
Sports & Outdoors	210	Teeth Whitening Products	19
Patio, Lawn & Garden	163	Power Dental Flossers	18
Electronics	134	Children's Electric Toothbrushes	13

Figure 1: Distribution of Product Prices

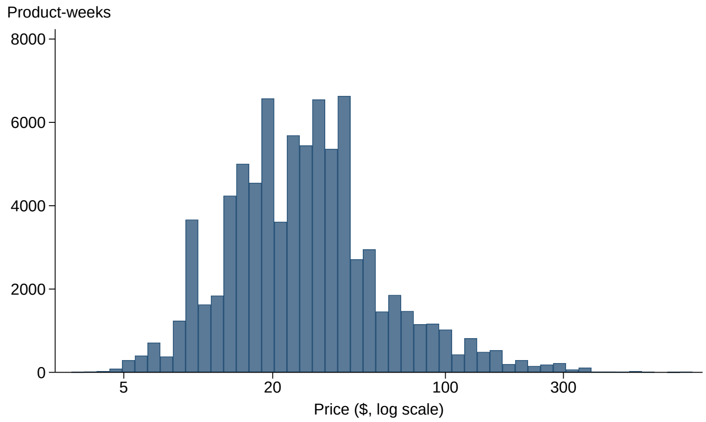


Figure 2: Average rating

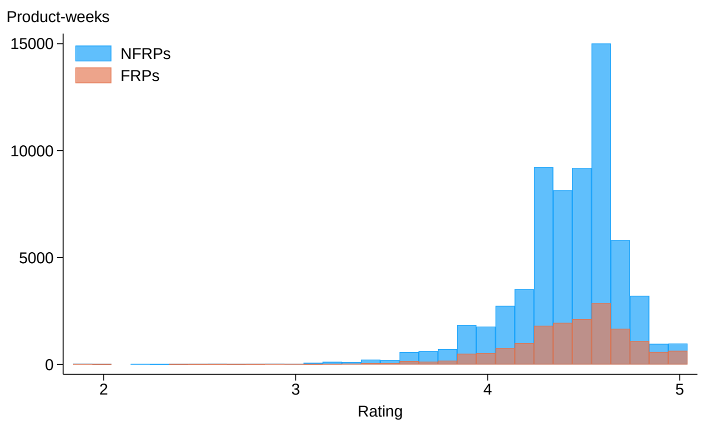


Figure 1 shows the distribution of product prices for the set of products observed buying fake reviews, which we refer to as the "focal products" or "fake review products" (FRPs). Most are under \$50 with a median price of \$24. Figure 2 shows the distribution of the products' average ratings, separately based on whether the product is an FRP or a "non-fake review purchaser" (NFRP). Most products have average ratings between 4 and 5 stars, with the focal products' ratings being inflated partially by fake reviews. Table 2 shows a full set of descriptive statistics on the focal and competitor products.

Sales Data For the demand model, it is necessary to have a measure of product-level market shares. Amazon does not report sales data directly, instead reporting a measure

Table 2: Characteristics of Focal Products and Comparison Products

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

called Best Seller Ranking or sales rank, which ranks products based on their rate of sales relative to other products in the same category. We collect sales rank for all products in our data on daily basis.

To calculate actual sales quantities, we exploit a feature of Amazon that makes product inventories observable for products with fewer than 1000 units in inventory. We collect this inventory data every 2 days for every focal and competitor product and use the changes over time in inventories to construct a measure of daily sales. For observations where this data is not available, we estimate a model relating sales to sales rank that fits the data well in-sample. This data and the model are described in detail in He and Hollenbeck (2020).

2.1 Estimating the Frequency of Fake Reviews

While we directly observe which products use fake reviews, we cannot know for sure which reviews are fake. Even while products are observed actively buying fake reviews, some share of the reviews they receive are likely organic. We will find it useful in our empirical model below, however, to estimate at the product level what share of reviews are fake. To do so, we rely on the insight found in He et al. (2022a), that products buying fake reviews must rely on a relatively small set of common reviewers participating in the Facebook groups. Therefore, products that share reviewers to an unusual degree are more likely to be rating manipulators.

He et al. (2022a) details a way to discern which products are fake review purchasers, given the network structure of reviews. They train a classifier model on features derived from the product-reviewer network as well as review features, text and photo features, and product metadata. This method performs well out of sample, detecting fake review buyers with an accuracy rate of .86 and AUC score of .93.

We use this prediction algorithm from He et al. (2022a) to classify all products in the product-reviewer network as buying fake reviews or not. This network is composed of all the FRPs and their competitors, as well as any other products that reviewers of these products also left reviews for. This consists of 25,840 products and 1.7 million reviews. We examine all the reviews of these products and identify any reviewers observed leaving multiple five-star reviews for products classified as purchasing fake reviews. We label these reviewers as "fake reviewers" and find 27,045 fake reviewers out of the 368,386 unique reviewers in this data, or roughly 7%. Then, for each product j that we know purchases fake reviews, we can compute the fraction of j's five-star reviews that came from these fake reviewers. This provides an estimate of θ_j^F as well as the overall distribution of θ_j^F as reported in Figure 19. For the products we observe buying fake reviews, the average estimated share of fake reviews is 56%

with a median share of 59%.[7]

3 A Simple Model of Misinformation and Mistrust

In this section, we present a simple model of how rating manipulation can affect market outcomes through two key channels: misinformation and mistrust. The "misinformation effect" of rating manipulation is that fake reviews provide false information that can mislead consumers into making different purchasing decisions. This is the direct effect that purchasing fake reviews has on a product's sales and the sales of its competitors. The "mistrust effect" is that consumer behavior will change if consumers are aware of fake reviews but cannot identify them. Consumers aware of fake reviews will account for the possibility that a product may have purchased fake reviews when interpreting its ratings. Mistrust is a more systemic effect, determined by the overall prevalence of fake review purchasing and not the specific purchasing of any individual product. Moreover, it affects products similarly regardless of whether that product purchased fake reviews. Indeed, the effect of mistrust can be felt even in markets where no products purchased fake reviews. Finally, while misinformation and mistrust represent effects on consumers' behavior, it is important to note that both also affect the equilibrium pricing behavior of both purchasers and non-purchasers of fake reviews.

3.1 Misinformation

To isolate the misinformation effect, we examine the effect of fake reviews on demand in a market in which consumers are unaware that some products may be purchasing fake reviews.

We model consumers' utility from a product j as decreasing in price (p_j) and increasing quality (q_j). However, when making purchasing decisions, consumers do not directly observe a product's quality and must infer it from the product's reviews (R_j). In our empirical

[7]By contrast, among non-fake review purchasing products, we observe only .6% of their reviews are left by these fake reviewers.

exercise, we think of R_j as a set of reviews that imperfectly reveal a product's quality. However, for simplicity in this toy model, we let R_j be a scalar rating that aggregates all of j's reviews and perfectly reflects j's true quality when j does not purchase fake reviews. Formally, we let $q_j, R_j \in (0,1)$ and $q_j = R_j$ when j does not purchase fake reviews. On the other hand, if a product purchases fake reviews, then $R_j \geq q_j$, and the ratings no longer perfectly reflect the true quality. We denote j purchasing fake reviews by F_j and j not purchasing fake reviews by $\neg F_j$.

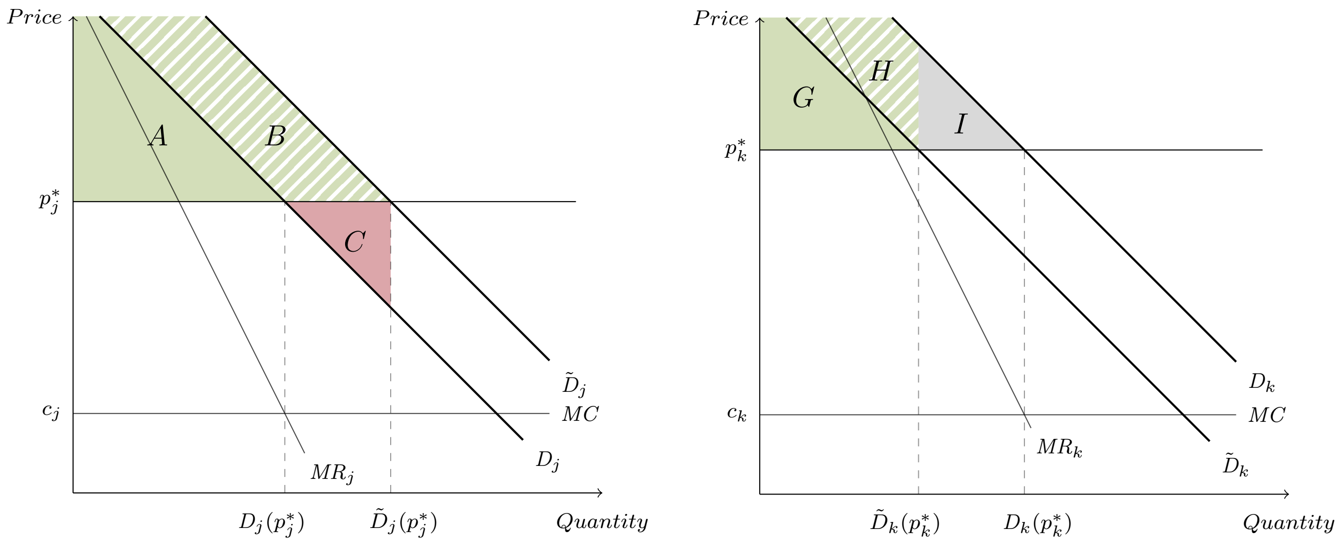
In a world without fake reviews, consumers perceive the relationship between ratings and expected quality as $E^T\left[q_j|R_j\right]$, where the superscript T indicates that consumer who trusts that no products purchase fake reviews. While E^T can be quite general, the relationship is simple given our assumptions: $E^T\left[q_j|R_j\right] = R_j$. In order to isolate the effect of misinformation, we fix consumers' perception of the relationship between ratings and quality to be trusting even when some products have purchased fake reviews.

For simplicity, we consider a market with two competing products, j and k. Given perfect information, the demand for product j—which we denote $D_j(p_j, q_j, p_k, q_k)$—is a function of j's own price and quality, as well as those of its competitor k. Since q_j and q_k are not observed, trusting consumers make their purchases based on observed ratings. We assume risk neutrality, so demand is then characterized by $D_j(p_j, R_j, p_k, R_k)$, where $R_j = E^T\left[q_j|R_j\right]$ and $R_k = E^T\left[q_k|R_k\right]$.

If product j purchases fake reviews, then this increases R_j above q_j and shifts out the demand curve for product j and shifts in the demand curve for competitor product k. Figure 3 shows the effect of these demand shifts when holding prices fixed. The demand curves D_j and D_k are those that would occur without fake reviews—i.e. with knowledge of q_j and q_k—and $\tilde{D}_j$ and $\tilde{D}_k$ are the demand curves given that j purchases fake reviews. Put simply, fake reviews cause consumers to purchase according to demand curves $\tilde{D}_j$ and $\tilde{D}_k$ even though the utility they actually receive from their purchases are determined by D_j and D_k.

For product j, this entails an increase in quantity demanded from $D_j(p_j^*)$ to $\tilde{D}_j(p_j^*)$,

Figure 3: Misinformation Effect (No Price Changes)



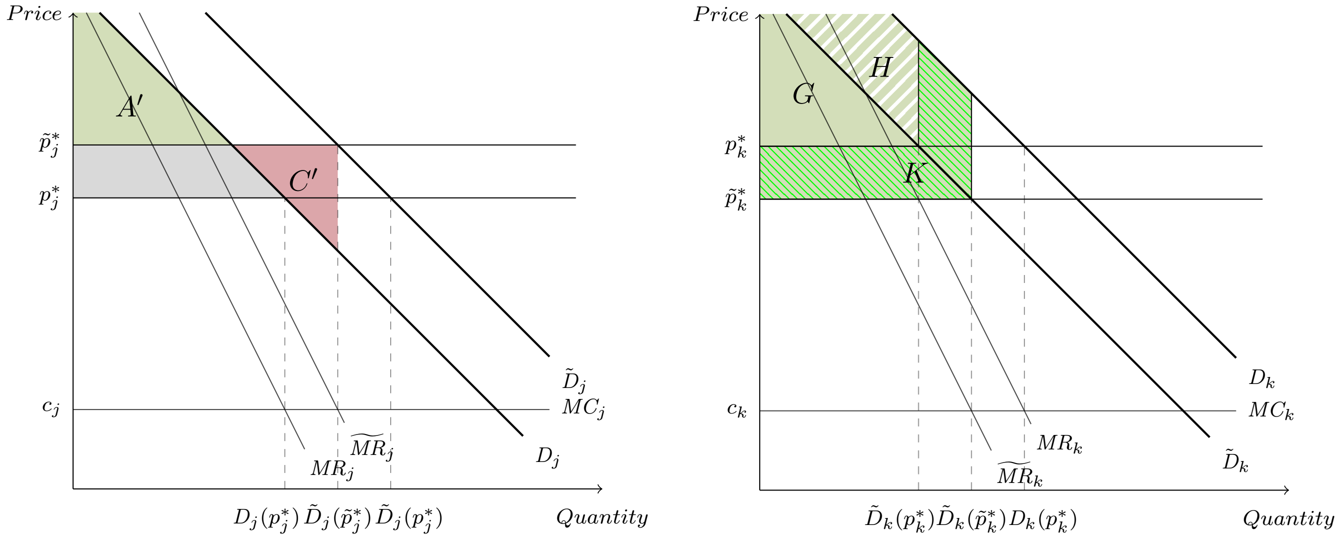
(a) Fake Review Purchaser

(b) Non-Purchaser

increasing j's profits by $\left(p_j^* - c_j\right)\left(\tilde{D}(p_j^*) - D(p_j^*)\right)$. Consumers purchasing based on $\tilde{D}_j$ anticipate a total consumer surplus of $A + B$. In actuality, however, consumer surplus for those purchasing j is much lower at $A - C$. Note that while fake reviews cause all consumers to overestimate the utility of purchasing j, not all purchasers of j are actually harmed. In particular, the $D_j(p_j^*)$ consumers who would have purchased j even absent fake reviews. For these consumers, region B only represents a failure of j to meet expectations and not an actual loss in utility. The true harms are borne by the $\tilde{D}_j(p_j^*) - D_j(p_j^*)$ consumers induced to purchase product j by its fake reviews. These consumers would have been better off either purchasing k or nothing at all, and the region C represents lost utility from making a sub-optimal purchasing decision due to misinformation.

The reduction in demand from $D(p_k^*)$ to $\tilde{D}_k(p_k^*)$ reduces k's profits by $(p_k^* - c_k)\left(D_k(p_k^*) - \tilde{D}_k(p_k^*)\right)$. Consumers purchasing based on $\tilde{D}_k$ anticipate receiving consumer surplus G. However, because the true utility from j is lower than consumers expected, those that purchased k are relatively better off than anticipated, receiving $G + H$. Of course, $\tilde{D}_k(p_k^*)$ consumers would have purchased k whether or not j purchased fake reviews, so H does not represent a real benefit. On the other hand, the $D_k(p_k^*) - \tilde{D}_k(p_k^*)$ consumers who are induced by fake reviews to purchase j instead of k are truly harmed.[8]

Figure 4: Misinformation Effect of j Purchasing Fake Reviews (With Price Changes)



(a) Fake Review Purchaser

(b) Non-Purchaser

Competitive Responses Previously, we held prices fixed at the levels that would have prevailed without fake reviews. Of course, j and k should adjust their prices in response to the shifts in demand. Figure 4a depicts these competitive responses to j purchasing fake reviews.

The increase in demand from D_j to $\tilde{D}_j$ raises j's optimal price from p_j^* to $\tilde{p}_j^*$. This increases j's profit to $\left(\tilde{p}_j^* - c_j\right) \tilde{D}_j(\tilde{p}_j^*)$ and shrinks consumer surplus from $A - C$ to $A' - C'$.[9] This induces an additional welfare loss with size $A - A'$. Note that this welfare loss impacts consumers who would have purchased product j even without fake reviews, but who now must pay higher prices as a result.

On the other hand, the decrease in demand from D_k to $\tilde{D}_k$ decreases k's optimal price from p_k^* to $\tilde{p}_k^*$. This increases k's profits to $(\tilde{p}_k^* - c_k)\, \tilde{D}_k(\tilde{p}_k^*)$ and increases consumer surplus to $G + H + K$. In this example, the $D_k(\tilde{p}_k^*)$ purchasers of k are actually better off than they

[8]Note that if fake reviews only steal market share and do not expand total purchasing in the market, then C and I represent the same harms due to misinformation.

[9]In this example with linear demand and fake reviews shifting only the level of demand, $C' = C$. However, this need not be the case in general.

would have been if j had not purchased fake reviews. That is, those consumers not induced to switch to product j benefit from the increased price competition.

3.2 Mistrust

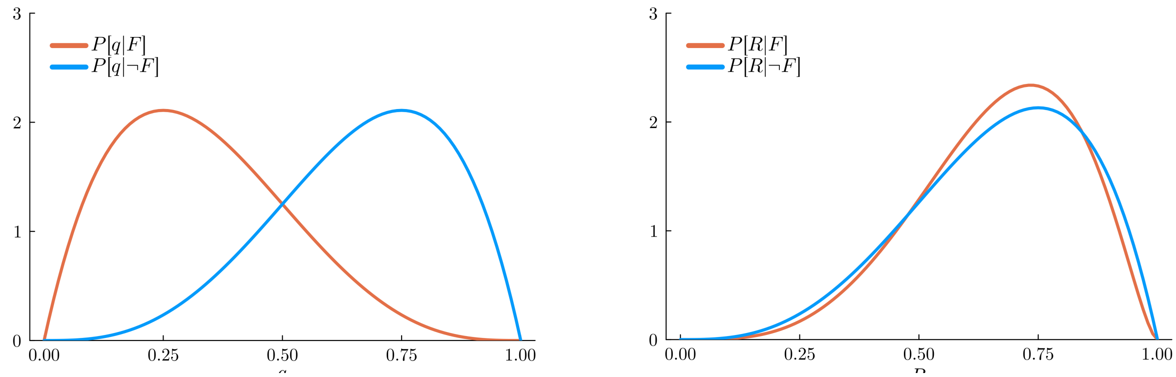
In order to isolate the effect of misinformation, we fixed consumers' perception of the relationship between reviews and quality to be $E^T[q|R] = R$, i.e. the relationship that would prevail in the absence of fake reviews. In this section, we examine the implications of consumers changing the way they interpret ratings in response to the general existence of fake reviews. Note that we largely suppress product subscripts in this section in order to emphasize that the effect of mistrust regards consumers' beliefs and not a given product's behavior. Indeed, mistrust may affect a market even if none of the products in that specific market purchase fake reviews so long as consumers believe that some products could be doing so.

We denote these updated beliefs about the relationship between ratings and quality by $E^M[q|R]$, where the superscript M indicates that consumers are mistrustful of ratings. We model this mistrust as consumers forming expectations about the quality of a product based on its ratings under the belief that the product may have purchased some fake reviews. In such a case, the consumer expects the quality to be between a convex combination of the expected qualities if a product did and didn't purchase fake reviews:

$$E^M[q|R] = P^M(F|R)\, E^M[q|R,F] + \big(1 - P^M(F|R)\big)\, E^M[q|R,\neg F]\,. \tag{1}$$

Equation (1) shows that the weights placed on each expectation are determined by $P^M(F|R)$, the perceived probability that a product with ratings R_j purchased fake reviews. Since we assume that absent fake reviews, $R = q$, rational consumers believe $E^M[q|R,\neg F] = R$. Furthermore, since we assume fake reviews are positive, rational consumers should perceive $E^M[q|R,F] < R$.

Figure 5: Toy Example Quality and Ratings



(a) Quality

(b) Ratings

Notes. $q|F$ is distributed Beta with mean $\frac{1}{3}$, and $q|\neg F$ is distributed Beta with mean $\frac{2}{3}$. For a fake review purchaser with quality q_j, the boost to their ratings due to fake reviews is $(1-q)\nu$ where $E[\nu] = 0.5$. Appendix section 9.2 details the relationship between q and R.

Figures 5 and 6 provide an illustrative example in which 50% of products purchase fake reviews and those that do tend to have lower quality than those that don't (Figure 5a). In this example, we assume that fake review purchasers increase their reviews from $R = q$ to $R = q + (1-q)\nu$, where ν is Beta distributed with mean 0.5 and standard deviation 0.2 (Appendix Figure 18). The result is that the distribution of ratings for products that purchase fake reviews is fairly similar to the ratings for products that don't (Figure 5b).

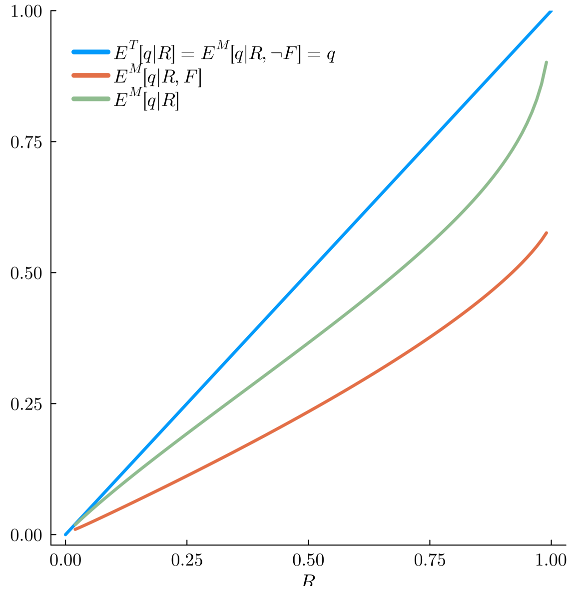


Figure 6: Expected quality.

Intuitively, the challenge that mistrusting consumers faces is that they are uncertain whether the product truly has quality R or whether R has been inflated through fake reviews. Figure 6 depicts how a Bayesian consumer infers quality from R given rational expectations about the prevalence of fake reviews and the joint distribution of q and R. (Note that we relax this assumption of rational expectations in our empirical exercise.[10]) The top line shows the quality that the consumer would infer if she knew that the product did not purchase fake reviews. This is simply q. The bottom curve gives the quality that the Bayesian consumer would infer if she knew that the product purchased fake reviews. The middle curve gives $E^M[q|R]$, the quality that the mistrusting consumer infers when observing rating R.

There are a number of instructive features of Figure 6. The first is that $E^M[q|R] \leq R = E^T[q|R]$, so mistrust causes consumers to anticipate lower utility from purchasing any product. This makes the option not to purchase any product more attractive and should increase the share of consumers choosing to do so. Mistrust can also result in substitution

[10]There are a number of reasons that consumers' beliefs about fake reviews may not satisfy rational expectations. For example, consumers may under- or overestimate the prevalence of fake reviews so that $P^M(F|R)$ is inaccurate. Or, they may misunderstand the extent to which fake reviews improve R so that $E^M[q|R, F]$ is inaccurate.

between products, which will depend on the own- and cross-quality elasticities of demand, as well as the relative extent to which mistrust reduces inferred quality. Figure 6 suggests the last can vary substantially with R. The reduction in expected quality from mistrust is largest when R suggests that fake reviews are either particularly likely—i.e. $P(F|R)$ is large—or that products achieving that R through manipulation are particularly bad—i.e. $E^M[q|R,F]$ is low.

In addition to shifting the level of perceived quality based on ratings, mistrust also changes the slope of perceived quality with respect to ratings. In turn, this implies that mistrust changes the quality elasticity of demand. In the example above, mistrust weakens the relationship between ratings and quality for most values of R. This is quite intuitive since it seems reasonable that consumers should be less responsive to a less trustworthy rating system. However, there can also be regions of R for which mistrustful consumers' are more responsive to increases in R than trusting ones.[11] In our example, this is true for high values of R, where an increase in R rapidly makes fake reviews both less likely and less likely to have dramatically affected R.

In sum, even this toy example suggests that the implication of mistrust for substitution patterns and quality elasticity of demand are highly dependent on many empirical factors, including the shape of consumer demand, the prevalence and magnitude of fake reviews, and the distribution of quality for both purchasers and non-purchasers. This complexity underscores the importance of the empirical exercise that we explore in the remainder of our paper.

It is important to re-emphasize that the scope of the effect of mistrust may be particularly large because it affects both products that did and did not purchase fake reviews similarly. In fact, it can affect markets in which no products actually purchased fake reviews as long as consumers perceive some probability that they could have. They are also difficult to

[11]Indeed, there are surprisingly few theoretical guarantees on the slope of $E^M[q|R]$. In fact, it it is straightforward to construct examples for which $E^M[q|R]$ is decreasing in R on some region. For example, this is true if modifying our toy example so that $R = \frac{1+q}{2}$.

measure or directly observe since they stem from consumers' perceptions. Finally, they may be difficult to attribute to individual actors, since the change in consumers' beliefs about the relationship between ratings and quality stems from the general prevalence of fake reviews and is not meaningfully shifted by the individual decisions of any single product.

4 Empirical Model

So far, we have modeled misinformation and mistrust in quite general terms. To make things more concrete, we must first precisely specify a model of how consumers interpret the ratings they observe. Section 4.1 presents a simple model in which Bayesian consumers observe the number of positive and negative reviews for each product and infer the product's quality under the assumption that reviews are independent and the probability that a given review is positive increases with the product's quality and if the seller purchased fake reviews.

This model suggests a few key components that we must estimate or assume. The first is consumers' priors about the distribution of product quality for products that do and do not purchase fake reviews. We estimate these in Section 4.2. The second is consumers' perceptions about the prevalence of fake reviews, which we estimate using an incentivized experiment in Section 5.

4.1 Consumer's Beliefs About Quality Given Ratings

In this section, we describe our model of how a Bayesian consumer forms beliefs about product quality based on observed ratings. Because the consumer is Bayesian, this entails detailing the assumptions the consumer makes about how reviews are generated.

We model consumers as considering each review r for a product as an independent signal of the product's quality. If the review is organic (i.e., not fake), then r is determined stochastically by the product's true quality q. For simplicity, we let reviews be either positive ($r = 1$) or negative ($r = 0$) with quality $q \in [0, 1]$ being the probability that an organic review is favorable. If all reviews were organic, the number of positive reviews that a given product receives out of N total reviews would be $B(N, q)$, i.e. binomial with success probability q.

Of course, not all reviews are organic. Some products purchase fake reviews, which we use indicator F to denote. If the product purchase fake reviews (i.e., if $F = 1$), then each review has $\theta^F \in (0, 1)$ probability of being fake. Taking this into account, the probability of a review being positive for a given product with quality q and fake-review purchasing behavior F is:

$$p_{Fq} := P(r = 1|q, F) = \begin{cases} q & \text{if } F = 0 \\ \theta^F + (1 - \theta^F)q & \text{if } F = 1. \end{cases} \tag{2}$$

Then, using this probability p_{Fq}, the split of N reviews between N^- negative and N^+ positive reviews is binomial $B(N, p_{Fq})$:

$$P(N^+|q, N, F) = \binom{N}{N^+} p_{Fq}^{N^+} (1 - p_{Fq})^{N^-}. \tag{3}$$

Given this, the consumer's posterior belief about the quality of a product with N^+ positive and N^- negative ratings is a straightforward application of Bayes' rule:

$$\begin{aligned} P\left(q \mid N^+, N\right) &= \sum_F P\left(F \mid N^+, N\right) P\left(q \mid N^+, N, F\right) \\ &= \sum_F P\left(F \mid N^+, N\right) \frac{P\left(N^+ \mid q, N, F\right) P\left(q \mid N, F\right)}{\int P\left(N^+ \mid q, N, F\right) dP\left(q \mid N, F\right)} \end{aligned} \tag{4}$$

Crucially, Equation (4) suggests that a few key terms required for our model. The first is $P(N^+|q, N, F)$, the probability of receiving N^+ positive reviews out of N reviews conditional on the product's quality and whether the seller purchases fake reviews. This term is binomial from (3). The second is $P\left(q \mid N, F\right)$, the latent distribution of quality for fake review purchasers and non-purchasers, which we estimate in Section 4.2. The third is $P\left(F \mid N^+, N\right)$, the consumer's perceived probability that a seller whose product has N^+ positive reviews out of N reviews is purchasing fake reviews. The last is θ^F, the consumer's perceived fraction of reviews that are fake for products that purchase fake reviews. These final two specifically regard consumer's perceptions on the prevalence of fake reviews, which we estimate via experiments in Section 5.

Finally, what appears in consumers' indirect utility function is:

$$\mathbb{E}\left[q|N^+, N\right] := \int q dP\left(q|N^+, N\right). \tag{5}$$

4.2 Estimating the Distribution of Latent Quality

Our model of consumers' Bayesian inference about product quality (Section 4.1) requires consumers' priors about the distribution of product quality for products that do and do not purchase fake reviews. We assume that consumers have correct priors about these distributions but do not condition their prior on the number of product reviews. The former assumption allows us to represent consumers' priors with an econometric estimate of the distributions of quality. The latter is that consumers implicitly assume $P\left(q \mid N, F\right) = P\left(q \mid F\right)$.[12]

To estimate these priors, we fit a distribution to maximize the average log-likelihood of the observed organic ratings. To do this, we first leverage our inferences in Section 2.1 to identify the products that purchase fake reviews and the number of fake reviews that each one purchased. Knowing this, we can compute the number of organic positive reviews—i.e., the number of positive reviews after deleting fake reviews—which we denote by N'^+. Likewise, we denote the number of organic reviews as $N' := N'^+ + N^-$.

We denote by $P(q|F;\gamma)$ the parameterization of $P(q|F)$ by γ. In our primary specification, we let q be Beta distributed conditional on F. In other words, $\gamma = \{(\alpha_F, \beta_F)\}_F$ and $q|F \sim \text{Beta}(\alpha_F, \beta_F)$. See Appendix 9.3 for additional details.

[12]This assumption reduces the dimensionality when estimating the priors. Note that it does not imply that consumers ignore the number of reviews, as this N still plays a key role how the consumer updates their beliefs based on ratings in equation (4).

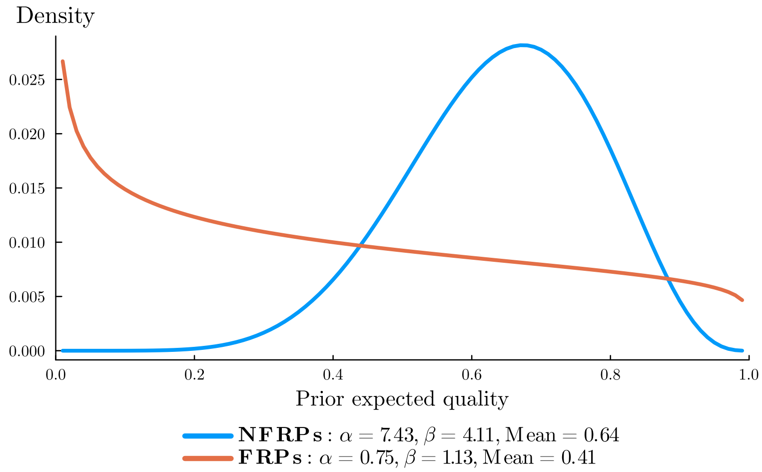


Figure 7: Estimated Priors

Using this, the likelihood of N^- negative and N'^+ organic positive ratings is:

$$LL\left(N^-, N'^+; \gamma\right) := \log\left(\int \binom{N'}{N'^+} q^{N'^+}(1-q)^{N^-} dP(q|F;\gamma)\right) \tag{6}$$

We estimate γ to be the maximizer of the log-likelihood of the organic reviews in the data:[13]

$$\hat{\gamma} := \arg\max_{\gamma} \sum_j LL\left(N_j^-, N'^+_j; \gamma\right), \tag{7}$$

where j indexes products in the data.

The estimated distributions for $P(q|F;\hat{\gamma})$ are shown in Figure 7. The estimates imply that products purchasing fake reviews tend to be of substantially lower quality than products that do not.[14] The average quality of a product that purchases fake reviews is 0.4, while the average quality of a product that does not is 0.64.

Note that even having estimated $P(q|F;\hat{\gamma})$, we still do not know the exact true quality of any individual product. However, because we can isolate organic reviews from fake reviews, we can use the estimates to infer a posterior distribution on quality for each product based

[13]Note that in practice the $\binom{N_j}{N_j^+}$ terms are additive and can be excluded as a constant.

[14]This finding is robust to alternative specifications, such as discretizing the unit interval and parameterizing $q|F$ to have a constant value on each sub-interval.

on only its organic ratings:[15]

$$
\begin{aligned}
P(q|N^-, N'^+, F; \hat{\gamma}) &= P(qN'^+|N', F) \\
&= \frac{P(N'^+|q, N', F)P(q|F; \hat{\gamma})}{P(N'^+|N', F)} \\
&= \frac{q^{N'^+}(1-q)^{N^-}P(q|F; \hat{\gamma})}{\int q^{N'^+}(1-q)^{N^-}dP(q|F; \hat{\gamma})}.
\end{aligned} \tag{8}
$$

We use these posteriors when computing the realized utility that consumers experience from their purchases.

5 Survey Experiments

This section describes a set of survey experiments run to help measure consumer beliefs about the prevalence of fake reviews. Our model of consumer beliefs about a product's expected quality takes the observed ratings distribution as inputs and computes the Bayesian posterior given some beliefs about fake review prevalence. The two values that enter the beliefs model are the probability that a given product j uses fake reviews $P(F_j|N)$, and the average fraction of reviews that are fake conditional on it doing so, θ_j^F.

These beliefs are inherently unobserved in our market-level data. The goal of these surveys, therefore, is to provide empirical grounding for the necessary assumptions we must make on these values. To do so, we focus on a set of prediction tasks designed to elicit these beliefs as well as how they vary across observable product features. We also ask a variety of direct questions about beliefs about fake reviews, their level of experience shopping on Amazon, and other demographic characteristics.

We observe the ground truth about which products use fake reviews and use this to make the survey payment incentive compatible. Each respondent's payout increases when they give correct predictions and this is communicated clearly to them (see details below).

[15] Where we have applied the assumption that $P(q|N', F) = P(q|F)$.

We also incorporate a reading comprehension check, an attention check, and an additional comprehension check for the main survey choice tasks in order to screen out bots or else humans who are not fully engaged with the survey. The survey is then run on Prolific, an online survey platform that connects researchers with a pool of potential survey respondents.

Prediction Task For each respondent, we begin by directly asking the question: "Out of 100 randomly chosen products on Amazon.com, how many would you expect to have purchased fake reviews?" Next, we show each respondent the 19 primary product categories on Amazon and ask them to select the 5 categories they most frequently shop in.

Respondents then move on to the main survey tasks. In these tasks, we show each respondent a set of 10 products, and for each, we ask their best guess as to the probability that that product uses fake reviews. The products they see are selected from the categories they selected previously. For each product, they are shown the product page as it appears on Amazon as shown in Figure 8 below. This displays the product name, image, price, average star rating, number of reviews, and other product details. Under the product page is a slider that asks "Using the slider below, please select the percentage probability on a scale of 0 to 100 that the product purchases or has purchased fake reviews.". When they engage the slider, it automatically updates and provides a full description of how their payout will vary depending on the answer they give and the underlying truth, as illustrated in Figure 9. By doing so, an engaged respondent will be fully informed that their optimal strategy is to reply with their true beliefs about the likelihood that the given product has used fake reviews. For each of the 10 probability questions, a respondent will receive \$1 if they answer "0%" to a product that does not use fake reviews or "100%" to a product that does. Each percentage point difference from the true value (0% or 100%) reduces their earning by 1 cent. Thus, for these 10 questions, a respondent can potentially earn a maximum of \$10, in addition to the base payment of \$1 that is paid to all respondents regardless of performance.

The products included in the survey are constructed as follows. First, 38 products (2

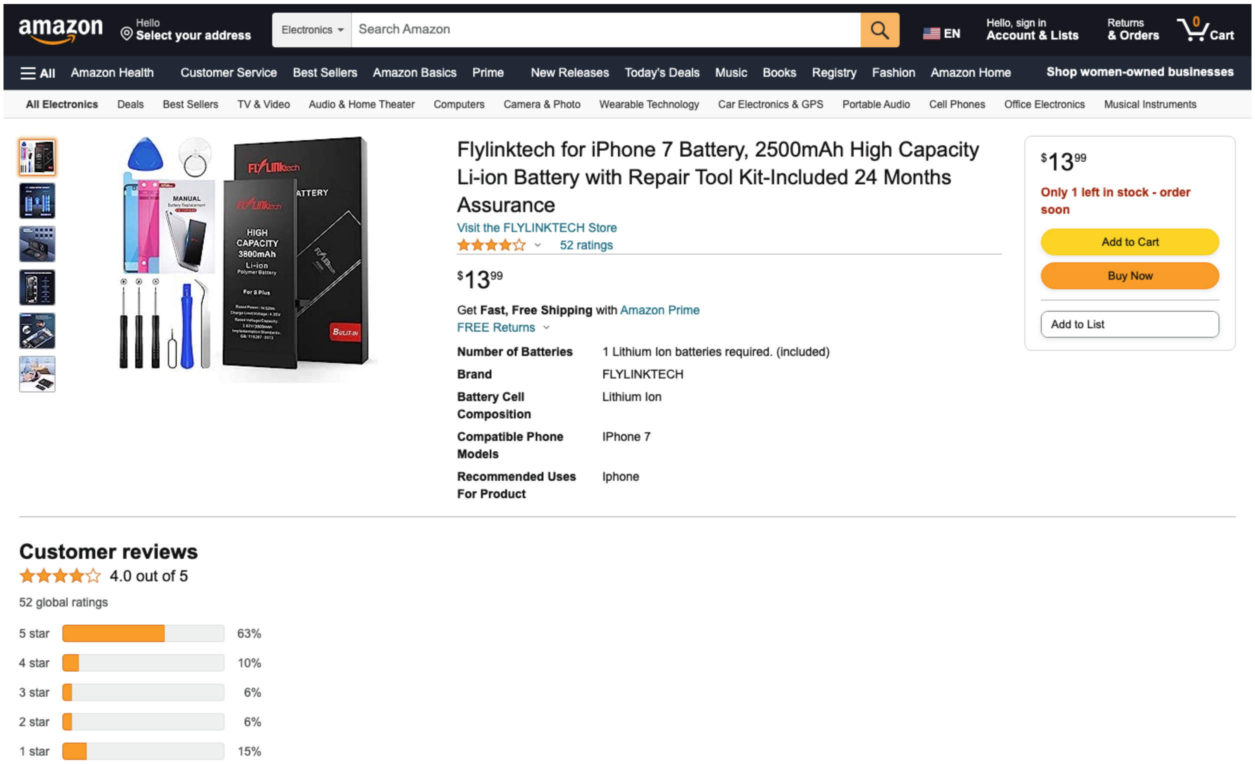


Figure 8: Example product page shown as in survey

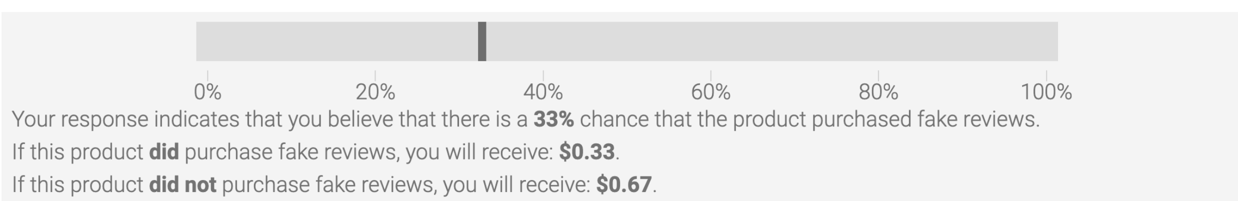


Figure 9: Slider after the respondent selects a probability.

from each category) are randomly drawn from a set of 1541 fake review purchasing products. Then, for each of these products, their closest competitor is included in the survey. For each question, the fake review purchaser is shown with a probability of 0.32, and the competitor is shown with a probability of 0.68, mimicking our estimated probability of seeing a fake review purchaser on Amazon. In addition, from our scraping of Amazon, we retain the underlying HTML file, allowing us to randomly vary certain product page components. We use this to generate additional random variation in the average rating and number of reviews displayed on the product page. This will allow us to test if beliefs vary with respect to these, holding the product itself constant. For each random draw of average rating and number of reviews, we alter the product rating histogram to match these by drawing the modal histogram from the underlying data among all products with those features. We also consider the possibility that even conditional on average rating and number of reviews, the shape of the histogram could affect beliefs about possible fake review activity. In particular, for the same average rating a product with all five-star and one-star reviews might be more suspicious than a product with a more uniform distribution. To capture this, we calculate the variance in ratings for each product in our data. Then, for 40% of respondents, we randomly show them not the modal histogram, but the histograms associated with the 5th or 95th percentiles of rating variance.

Finally, for some respondents, we include as a comprehension task an Amazon gift card as the product and assume that respondents would reasonably attach zero probability to the likelihood that this product is using fake reviews.

5.1 Survey Results

We ran an online survey in July 2023 and summarize the results here. Our final run produced a sample size of 401 qualified respondents who passed the reading comprehension and attention checks, out of an initial sample of 711. Their demographic characteristics are summarized in Table 3.

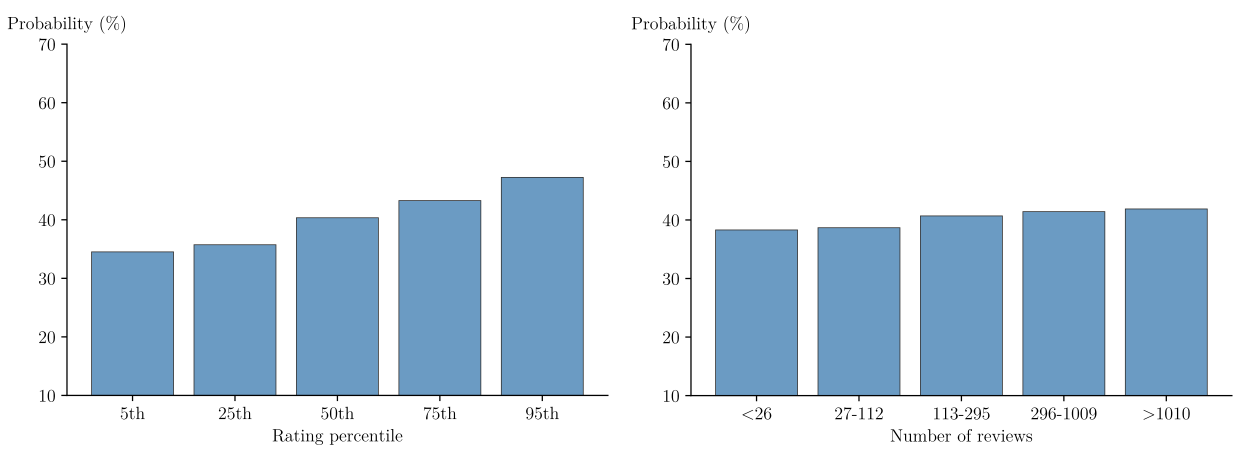
Table 3: **Demographics of Survey Respondents**

Age	
10-19	0.00
20-29	0.19
30-39	0.42
40-49	0.19
50-59	0.14
60-70	0.04
70+	0.02
Gender	
Male	0.56
Female	0.43
Non-binary/third gender	0.01
Household Size	
1	0.26
2	0.30
3	0.20
4	0.17
5	0.06
6+	0.01
Income ($)	
0-9999	0.06
10000-14999	0.04
15000-19999	0.02
20000-29999	0.11
30000-39999	0.15
40000-49999	0.07
50000-69999	0.20
70000-89999	0.13
90000-109999	0.10
110000-149999	0.06
150000-199999	0.04
200000+	0.02
Education	
Some high school or less	0.01
High school diploma or GED	0.13
Some college, but no degree	0.24
Associates or technical degree	0.09
Bachelor's degree	0.35
Graduate or professional degree	0.17
Prefer not to say	0.00

For the initial question, asking "Out of 100 randomly chosen products on Amazon.com, how many would you expect to have purchased fake reviews?" the mean response is 31% and the median is 26%. This is slightly lower than the 32% of products we observe in our data. For the prediction task questions, beliefs about fake review prevalence are somewhat higher. In instances where the respondent is shown a fake review product, the mean response is 42% and the median is 40%. In cases where the product shown does not use fake reviews, the mean is 39% (median 36%). The fact that their probabilities are so similar for the two product types suggests that consumers do a poor job of predicting which products use fake reviews based on viewing the product page.

We also examine how these probabilities vary with respect to the product's average rating and number of reviews. The results are shown in Figure 10, in the left panel we compare the mean probability for products with the 5th percentile of average ratings up to the 95th percentile. There is a clear upward trend, where products with very high ratings are seen as more likely to be using fake reviews than products with very low ratings. In the right panel, we show results for the number of reviews. There is no apparent relationship between the number of product reviews and consumer beliefs about the likelihood of fake review activity.

Figure 10: Beliefs About Fake Reviews by Product Characteristics



(a) By Rating

(b) By # Reviews

Next, we examine how the mean response varies across the full set of interactions between

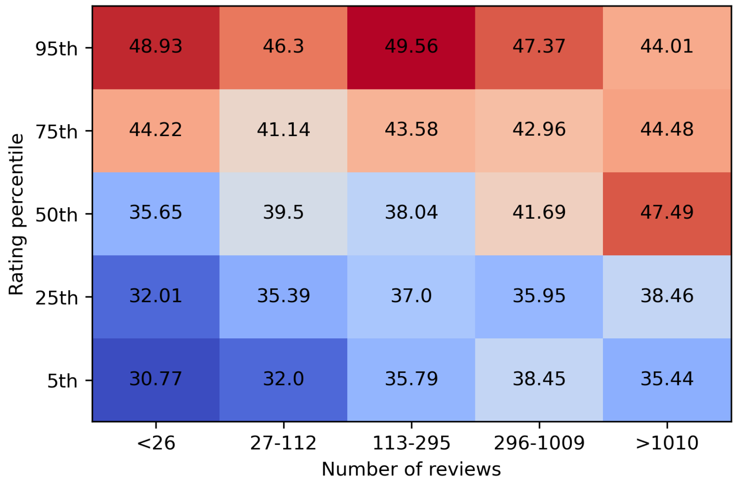


Figure 11: Beliefs by Product Characterstics

the average rating and the number of reviews. Figure 11 shows the full results in heatmap form. We see that consumers are especially suspicious of products with very few reviews but a very high average rating. Products with few reviews but low ratings, by contrast, have the lowest level of predicted fake review activity. For products with a large number of reviews, there is a substantially weaker relationship between the average rating and beliefs about fake review likelihood.

Finally, for the last of the ten products respondents are shown, we ask the follow-up question: "For this question, please assume that this product has purchased fake reviews. Guess the fraction of fake reviews among all its reviews." This is meant to elicit beliefs about θ^F, the proportion of fake reviews for products known to be using them. This task is also incentive compatible. We get a mean response of 38% and a median of 31%.

For the question that displays the Amazon gift card, 50% of the respondents correctly responded 0%, and the mean response is 11%. Figure 12 shows the histogram of responses. We test for a relationship between giving a response greater than 10% to the gift card question and other survey responses and find no relationship, and overall results are similar when this group are excluded.

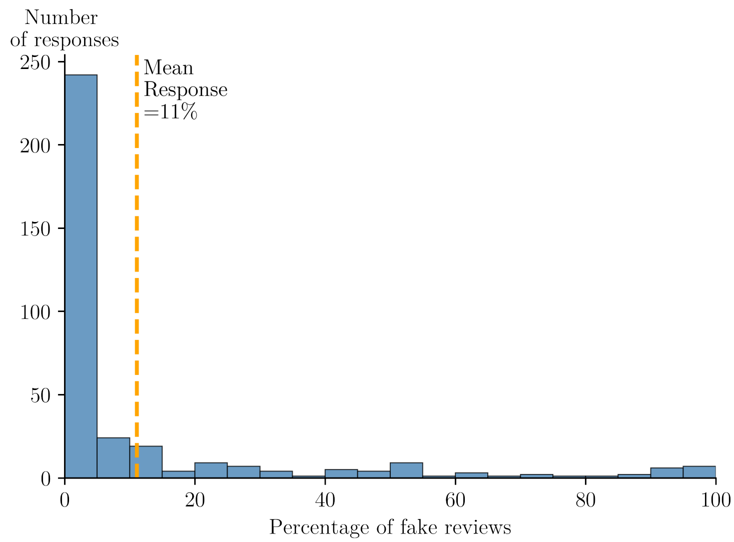


Figure 12: Responses for Amazon Gift Card

6 Consumer Demand and Welfare

In this section, we specify a model of consumer demand given ratings, prices and other attributes. We then describe how this model is identified and estimated and present estimation results.

6.1 Consumer Indirect Utility

We model demand using the standard discrete choice random utility framework following the approach of Berry et al. (1995). Consumers make a purchase decision based on their indirect utility function, specified as:

$$u_{ijt} = \beta_{0i} E(q^*_{jt} | r_{jt}, N_{jt}) - \alpha_i p_{jt} + \beta X_{jt} + \xi_j + \lambda_t + \epsilon_{ijt} \tag{9}$$

where $E(q^*_{jt})$ is the consumer's belief about quality given its star rating and number of reviews as described in the previous section. Price p_{jt}, product age (cumulative time listed on Amazon), and position in search results also enter into indirect utility, as do product fixed effects ξ_j and time fixed effects λ_t. We assume that consumers are not forward-looking or strategic in the timing of their purchases. To allow for heterogeneity in individual preferences, we model consumer utility over expected quality as $\beta_{0i} = \beta_0 + \nu_i$, where $\nu_i \sim log\mathcal{N}(0, \sigma)$. The use of a lognormal distribution of individual heterogeneity restricts preferences such that all consumers place positive weight on expected quality.

We define market at the keyword-week level and denote the set of products in the market as $\mathcal{J}$. To construct a manageable set of competitors, we choose the set of up to ten products that co-occur most frequently with each focal product in our keyword search results.

The mean value of the outside option of not purchasing or purchasing from a different platform is normalized to zero. We follow Grigolon and Verboven (2014) in modeling correlation in preferences over certain products, in this case, all inside good products in the same market. This allows for the possibility of more substitution between products within a subcategory than across and better captures substitution to the outside good.

Specifically, the idiosyncratic term $\bar{\epsilon}_{ijt}$ follows the nested logit distribution, where products in the same group have correlated preferences. We can, therefore, write this term as:

$$\bar{\epsilon}_{ijt} = \zeta_{igt} + (1-\rho)\epsilon_{ijt}, \tag{10}$$

where $\rho \in [0, 1]$ and represents a nesting parameter.

Denote the mean component of utility $\delta_{jt} = \beta_0 E(q^*_{jt}|r_{jt}, N_{jt}) - \alpha_i p_{jt} + \beta X_{jt} + \xi_j + \lambda_t$. This utility and the error structure just described generate the following conditional probability that consumer i purchases product j from market g:

$$s_{igjt}(\delta_{jt}, \theta, \nu_i, D_i) = M \cdot \frac{\exp((\delta_{jt})/(1-\rho))}{\exp(I_{igt}/(1-\rho))} \frac{\exp(I_{igt})}{\exp(I_{it})}, \tag{11}$$

where $\theta = (\beta, \alpha, \rho)$ and I_{igt} is an inclusive value term such that

$$I_{igt} = (1-\rho) \log \Sigma_{j \in G} \exp((\delta_{jt})/(1-\rho)) \tag{12}$$

$$I_{it} = \ln(1 + \Sigma_g \exp(I_{igt})). \tag{13}$$

Total weekly sales quantity equals this market share times time-varying market size $M_{\mathcal{J},t}$. We define $M_{\mathcal{J},t}$ by taking the moving average of total weekly sales for the products in $\mathcal{J}$ at the monthly level and multiplying by a constant. The parameters of this demand function are estimated using weekly data on market shares, ratings, number of reviews, and prices for all products in the consideration set. We describe this estimation next.

Beliefs Consumer utility is a function of beliefs about expected product quality given ratings $E(q^*_{jt}|r_{jt}, N_{jt})$. These beliefs are not identified from demand and so we do not estimate them jointly with the demand parameters. Instead, we use the model described in section **??**, in which a Bayesian consumer takes the observed average rating and number of reviews and forms expectations of product quality from these. This model relies on unobserved beliefs about the prevalence of fake reviews.

To implement the model and compute the expected quality given reviews, we incorporate our survey results as described in section 5.1 into the model. In particular, the two values needed are $P(F)$, the probability that a given product uses fake reviews, and θ^F, the average fraction of reviews that are fake if it does so. We compute these values as a function of both the number of reviews and average rating as described in section 5.1. That is, we allow beliefs about the likelihood a product uses fake reviews to differ for products with few reviews and high ratings, many reviews and high ratings, and so on. As a benchmark, we also test a rational expectations set of beliefs, in which the true average proportions in our data are used in place of these, although as noted in section 5.1 the beliefs elicited in our survey experiments are close to these true proportions.

6.2 Estimation and Identification

To estimate the model we use a GMM estimator that interacts the structural demand side error $\omega(\theta)$ with a set of instruments Z, where the demand parameters are $\theta = (\alpha, \beta, \sigma, \rho)$.

Formally the GMM estimator is formed from the population moment condition $E[Z' \cdot \omega(\theta)] = 0$. The GMM estimate is

$$\hat{\theta} = \min_{\theta} \omega(\theta)' Z A^{-1} Z' \omega(\theta) \tag{14}$$

for some positive definite weighting matrix A. To construct the structural error $\omega(\theta)$ we use the BLP contraction mapping to obtain the unique vector $\delta^*(x_{jt}, S_{jt}, \theta)$, which maps the observed market shares S_{jt} into mean utility values. A 2SLS regression of $\delta^*(x_{jt}, S_{jt}, \theta)$ on product characteristics, price and various fixed effects with instruments Z then produces a residual term that is equivalent to $\omega(\theta)$. In our 2-step GMM we use $A = Z'Z$ in the first step and in the second step construct the heteroskedasity robust optimal weighting matrix clustered at the retailer level. We implement the demand estimation using the pyblp package and following best practices as described by Conlon and Gortmaker (2020), which we find to converge consistently.

The first challenge in identification is in identifying price sensitivity α, as prices may be correlated with the unobserved product-level demand shocks ϵ_{ijt}. We follow a standard approach and use Gandhi and Houde (2019) instruments constructed from the product characteristics of competing products. Another concern is that the product ratings in $E(q^*_{jt}|r_{jt})$ are correlated with the unobserved product quality. We rely on product fixed effects to absorb mean product quality. Thus, we treat the variation in ratings over time as largely exogenous.

Lastly, we need additional instruments for the nesting parameter and so require instruments that generate variation in the conditional shares of the inside good. We use the number of products in the market, a standard instrument for this problem (see Miller and Weinberg (2017).)

6.3 Results of Demand Estimation

In Table 4 we show the results from demand estimation. Our preferred specification includes product and week fixed effects and uses Gandhi-Houde IVs for price as well as the number of products in the market.

We find the elasticity of demand with respect to expected product quality is fairly high at roughly 2. This is not directly comparable to previous estimates of the elasticity with respect to ratings. We find a mean price elasticity of -6.2 with a median of -4.3.

Table 4: Results of Demand Estimation

Price	-0.16
	(0.017)
$E(q\|N^+, N^-)$	0.88
	(0.059)
σ	5.2e-06
	(0.016)
Age	-0.02
	(0.026)
Listing Rank	-0.018
	(0.0027)
ρ	0.21
	(0.041)
Product FEs	Yes
Week FEs	Yes
Gandhi-Houde IVs	Yes
Median Own-Price Elast.	-4.3
Mean Own-Price Elast.	-6.2
Median Own-Quality Elast.	2
Mean Own-Quality Elast.	2
Observations	81,364

7 Counterfactuals

To measure the net effects of rating manipulation on firm outcomes and consumer welfare, we conduct a series of counterfactual analyses in which the platform credibly eliminates fake reviews, and both firms and consumers adjust their behavior. Implementing this analysis consists of several parts. First, we compare consumer beliefs about product quality, as well as prices and quantities sold and seller profits, between the factual world where fake reviews are present and consumers are mistrustful of reviews to the counterfactual world in which no fake reviews are present and consumers are fully trusting of reviews. Second, to isolate the misinformation and mistrust channels we evaluate separate counterfactuals in which Amazon eliminates fake reviews but consumers remain mistrustful and in which consumer mistrust is eliminated but fake reviews remain. In each case, we consider separately the role of competition in these changes by holding prices fixed vs allowing firms to react by changing prices.

7.1 Full Equilibrium Counterfactual

We start by recomputing product ratings after the elimination of fake reviews. We use the method described in section 2.1 to estimate the share of reviews that are fake for each product in our data. Given that all of these are five-star reviews, to simulate the platform deleting their fake reviews, we simply need to adjust their average rating and number of reviews downward based on the proportion of five-stars that were removed.

Figure 13: Average rating, fake review purchasers

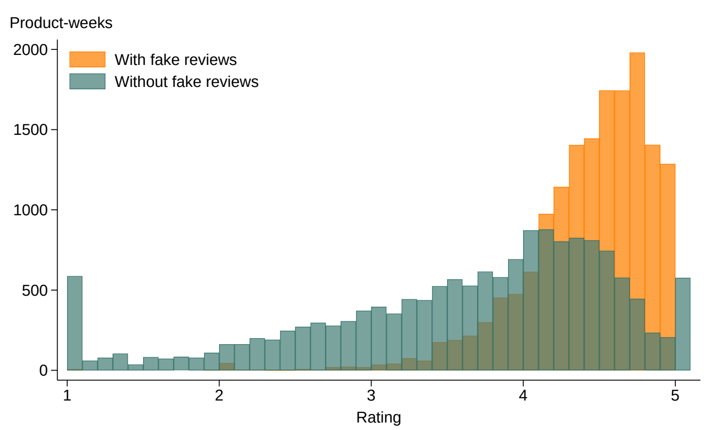
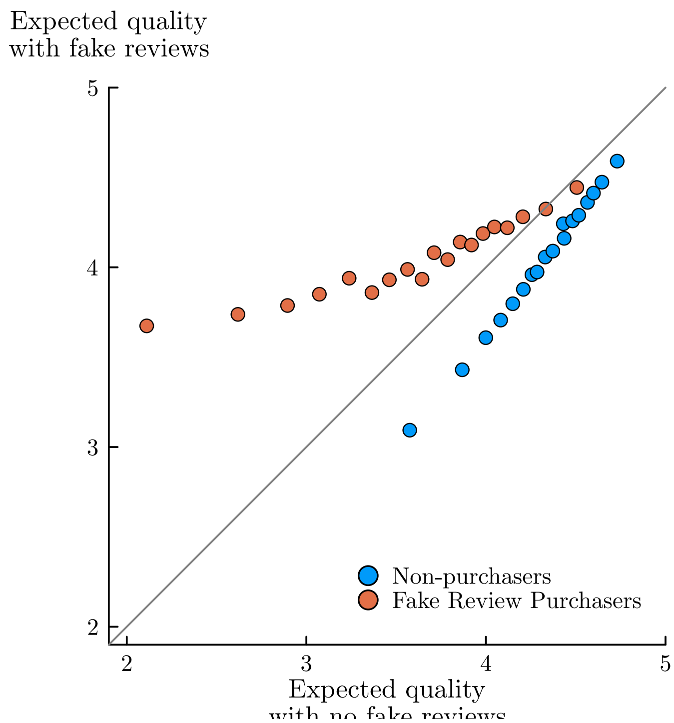


Figure 13 shows the distribution of average ratings for focal products when the fake reviews are included compared to when they are absent. Next, these average product ratings and numbers of reviews are used to compute expected quality. In Figure 14 the perceived qualities without fake reviews are plotted against the perceived qualities with fake reviews present. We show this separately for products that use fake reviews vs those that do not. For fake review products, their perceived quality with fake reviews increases substantially, particularly for products with relatively low ratings. For non-fake review products, the presence of fake reviews in consumer beliefs causes their perceived quality to fall as a result of consumer mistrust.

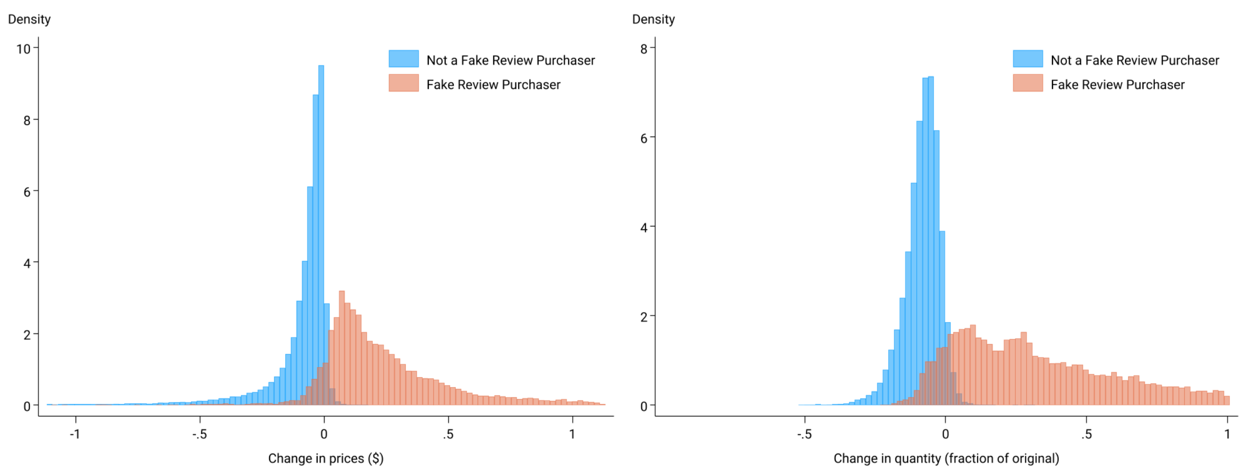
Figure 14: Perceived Product Quality with and without Fake reviews



Next, we solve for product demand under the factual and counterfactual set of perceived qualities and allow sellers to adjust their prices. We run the simulated outcomes for all weeks in our data sequentially and note that there is a potentially important feedback loop embedded in the data. If in period 1, an FRP loses sales in the counterfactual relative to an NFRP, in period 2, this could impact their relative positions in search rankings because these are a function of past sales. To account for this, we estimate a hedonic model of product position and in our counterfactual allow for period t sales to impact period $t+1$ positions.

The full set of results are summarized in Table 5, with the outcomes for the world without fake reviews shown in the first column and the outcomes for the full equilibrium with fake reviews in the rightmost column. Figure 15 visually depicts the changes in equilibrium prices with fake reviews present vs absent. The median change in prices is an increase of \$0.28 for fake review purchasers and a decrease of \$0.07 for non-purchasers. Fake review purchasers are able to charge higher prices because of the upward adjustment in their ratings, and because of consumer mistrust due to fake reviews, the other products decrease their prices. The net sales-weighted average price difference due to the presence of fake reviews is -\$0.09 and the net change in sales quantities is roughly -2%.

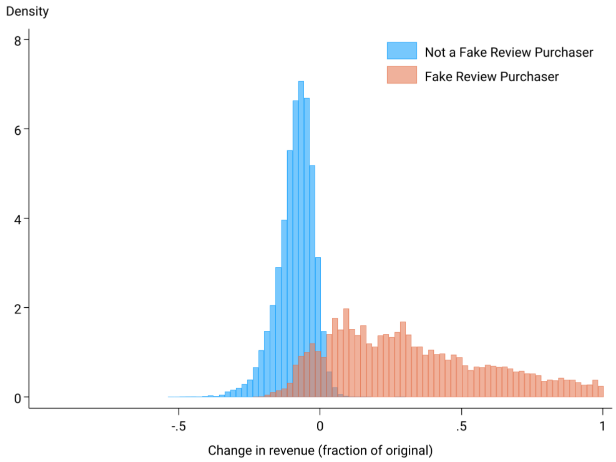
Figure 15: Counterfactual differences in prices and quantities

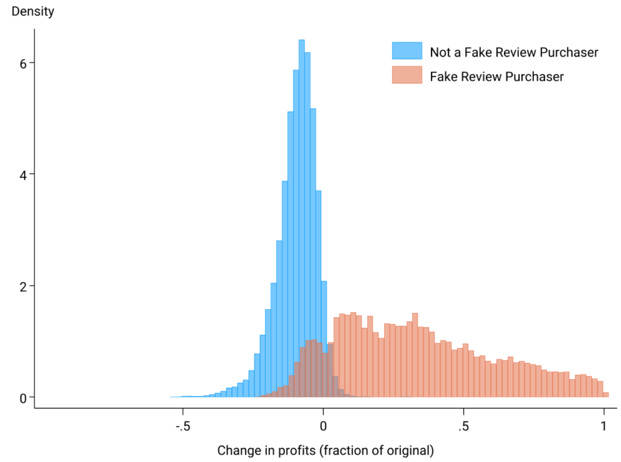


Next we calculate how these differences in prices and sales quantities translate into product level revenues and profits. The distribution of changes is shown in Figure 16. As expected, the presence of fake reviews causes revenue and profits to increase for those products using fake reviews and to fall for the others. The average net effect is an overall increase in both revenue and profits for sellers, and hence for Amazon who receives a fixed commission on these revenues.

Lastly, we compute the change in net consumer welfare due to the presence of fake reviews. We compute the realized utilities in both scenarios, which is not the same as

Figure 16: Counterfactual differences in revenues and profits





the expected utility at the time of purchase. In the equilibrium with fake reviews present, the expected quality that enters into the consumer's choice will be different from the true underlying product quality. Whereas market shares are determined by the expected quality, the realized utility should be calculated at the true quality. We therefore compute experience utility $\tilde{u}_{ijt}$ with an offset term that depends on the discrepancy between perceived and true qualities and the estimated coefficient on quality. This offset term is calculated as:

$$\Delta q := q_{perceived} - q_{true}$$

$$\tilde{u}_{ijt} = u_{ijt} - \beta_1 \Delta q_{ijt}$$

The welfare for consumer i in market t is then

$$\begin{aligned} W_{it} &= E_\epsilon[u_{ij^*t}] - E_\epsilon[\Delta q_{ij^*t}] \\ &= \bar{W}_{it} - \sum_{J_t} s_{ijt}(\beta_1 \Delta q_{ijt}), \end{aligned}$$

where j^* is chosen based on perceived quality, and $\bar{W}_{it}$ is the welfare evaluated using decision utility. For more on this adjustment, see Train (2015) or Reimers and Waldfogel (2021). We find a net welfare loss to consumers when fake reviews are present, where the magnitude of the welfare loss is roughly equivalent to 0.5% of the median product purchase price or a loss of about $.15 per consumer per week.

7.2 Misinformation and Mistrust Counterfactuals

Next, we compute the full set of market outcomes for counterfactual scenarios designed to isolate the impacts of misinformation and mistrust. In both cases, we first compute the results holding firm prices fixed and then allowing firms to adjust prices in order to also isolate the competitive responses to each mechanism.

We start from the baseline of a market without fake reviews and where consumers fully trust product ratings. We first isolate the misinformation effect by re-introducing the observed fake reviews for products that use them but keeping fixed consumers' beliefs about the relationship between ratings and quality. That is, consumers continue to trust reviews. Second, to incorporate the mistrust effect, we again start from the baseline of a market without fake reviews and introduce mistrust by allowing consumers to believe fake reviews exist without actually re-introducing the observed fake reviews. In other words, we isolate mistrust from misinformation by considering a counterfactual in which there are no fake reviews, but consumers mistrust ratings as if there were.

The results are shown in Table 5. We find that misinformation alone causes only a small decrease in consumer welfare. In this scenario consumers shift their purchases towards FRPs, who charge higher prices, but many still buy NFRPs whose prices have fallen. When we compare this to the effects of mistrust, we find that mistrust causes a much larger decrease in consumer welfare. Mistrust causes consumers to buy fewer of each type of products. Mistrust does cause prices to fall via increased competition, but the price effects are relatively small. On net, when both effects are present consumers are harmed and it is clear that a substantial majority of this harm results from lack of trust in ratings. We also find that price competition plays an important role in this. If prices were held fixed consumers would be substantially worse off, but the fall in prices for NFRPs offsets the increase in prices from FRPs enough to partially alleviate the welfare harms from fake reviews.

Notably, these counterfactuals also shed light on the platform's incentives. We calculate platform revenue as a fixed share of sales, using the actual platform commissions charged by Amazon. We find that the platform generally benefits when consumers do. A challenge, however, is that if Amazon simply deletes fake reviews without consumers adjusting their beliefs, platform revenue falls. That is, moving from the full equilibrium to the mistrust only scenario is harmful to platform profits, whereas if mistrust was also eliminated, they (and consumers) would be better off. From the platform's perspective, that is, simply removing fake reviews could backfire in the short run if consumers are not informed about this or do not find it credible.

Table 5: Outcomes in Counterfactuals

	No FR	Misinfo	Mistrust	Misinfo+Mistrust	
		Floating prices	Floating prices	Fixed prices	Floating prices
Welfare ($)	61,791,060	61,688,828	60,581,713	60,279,612	60,665,290
Platform revenue ($)	19,143,666	19,247,171	18,285,577	18,388,358	18,446,978
FRP average prices ($)	31.30	31.68	31.18	31.30	31.62
NFRP average prices ($)	37.86	37.79	37.82	37.86	37.75
FRP sales (units)	645,421	901,136	568,366	926,511	866,698
NFRP sales (units)	4,569,797	4,395,619	4,408,504	4,147,416	4,213,520
FRP profits ($)	4,412,089	6,269,425	3,843,358	6,057,776	5,979,472
NFRP profits ($)	28,575,985	27,136,755	27,425,096	25,814,701	25,850,417

8 Conclusion

A core mission of consumer protection regulators is to prevent firms from engaging in deceptive practices. A form of deception of growing importance is the manipulation of reputation systems by sellers on two-sided online platforms. In this paper we bring new empirical evidence on the magnitude and nature of consumer harms from this practice in a highly relevant empirical seeing, the use of fake product reviews by third-party sellers on Amazon.com.

There are two channels by which rating manipulation impacts consumer welfare. The first is the direct effect of the deception, which we refer to as misinformation. Ratings inflated by fake reviews shift demand from high-quality to low-quality products and allow low-quality sellers to charge higher prices. The second is the indirect effect on consumer perceptions of the trustworthiness of ratings. These effects are more ambiguous, as low trust in ratings may cause consumers to make worse purchase decisions but they may also increase price compensation sufficiently to offset this.

We formalize these effects by explicitly deriving a model of how consumers form expectations of product quality from ratings as well as how these beliefs are contingent on beliefs about the trustworthiness of ratings. This model of beliefs is then incorporated into a model of product choices and utility. We evaluate this model empirically using a novel dataset on several thousand products on Amazon for which can directly observe their fake review activity. By estimating our model from these data, we can then simulate the removal of fake reviews and quantify the different channels by which consumers are impacted by rating manipulation.

We find that the presence of fake reviews both makes consumers worse off and reduces platform profits. However, when consumers are fully informed that fake reviews exist and adjust their trust in ratings accordingly, the increase in price competition substantial ameliorates welfare losses. This highlights the importance of beliefs and equilibrium price responses in this result.

References

Akesson, J., Hahn, R. W., Metcalfe, R. D., and Monti-Nussbaum, M. (2022). The impact of fake reviews on demand and welfare. *Working Paper.*

Armstrong, M. and Zhou, J. (2022). Consumer information and the limits to competition. *American Economic Review*, 112(2):534–77.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63:841–890.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Chakraborty, I., Kim, M., and Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, 59(3):600–622.

Commission, F. T. (2023). Trade regulation rule on the use of consumer reviews and testimonials. 16 CFR 465: 88 FR 49364, RIN: 3084-AB76.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. *RAND Journal of Economics.*

Dai, W. D., Jin, G., Lee, J., and Luca, M. (2018). Aggregation of consumer ratings: an application to Yelp.com. *Quantitative Marketing and Economics (QME)*, 16(3):289–339.

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

Dranove, D. and Jin, G. Z. (2010). Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*, 48(4):935–963.

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

Gandhi, A. and Houde, J.-F. (2019). Measuring Substitution Patterns in Diï¬*erentiated* − *ProductsIndustries.NBERWorkingPapers*26375, *NationalBureauofEconomicResearch*, *Inc.*

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

Grigolon, L. and Verboven, F. (2014). Nested logit or random coefficients logit? a comparison of alternative discrete choice models of product differentiation. *The Review of Economics and Statistics*, 96(5):916–935.

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

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., and Tosyali, A. (2022a). Detecting Fake Review Buyers Using Network Structure: Direct Evidence from Amazon. *Proceedings of the National Academy of Sciences*, 119(47).

He, S., Hollenbeck, B., and Proserpio, D. (2022b). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Hopenhayn, H. and Saeedi, M. (2023). Optimal Information Disclosure and Market Outcomes. *Theoretical Economics*.

Hui, X., Jin, G. Z., and Liu, M. (2022). Designing Quality Certificates: Insights from eBay. NBER Working Papers 29674, National Bureau of Economic Research, Inc.

Hui, X., Saeedi, M., Shen, Z., and Sundaresan, N. (2016). Reputation and regulations: Evidence from ebay. *Management Science*, 62.

Li, L. I., Tadelis, S., and Zhou, X. (2020). Buying reputation as a signal of quality: Evidence from an online marketplace. *RAND Journal of Economics*, 51(4):965–988.

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.

Miller, N. and Weinberg, M. (2017). Understanding the Price Effects of the MillerCoors Joint Venture. *Econometrica*.

Reimers, I. and Waldfogel, J. (2021). Digitization and Pre-purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings. *American Economic Review*, 111(6):1944–1971.

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

Saeedi, M. and Shourideh, A. (2020). Optimal Rating Design under Moral Hazard. Papers 2008.09529, arXiv.org.

Saraiva, G. (2020). Incentives to fake reviews in online platforms.

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

Train, K. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of choice modelling*, 16(C):15–22.

Vatter, B. (2021). Quality disclosure and regulation: Scoring design in medicare advantage.

Vellodi, N. (2018). Ratings design and barriers to entry. Working Papers 18-13, NET Institute.

Wu, C., Che, H., Chan, T. Y., and Lu, X. (2015). The economic value of online reviews. *Marketing Science*, 34(5):739–754.

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

9 Appendix

9.1 Computing welfare

The consumer's expectation of product quality is in general different from the econometrician's estimate of latent quality, since the consumer has less information about fake reviews. For the purchasing decision, the relevant quantities are the consumer's perceived quality, which enter the decision utility; for welfare, the relevant quantities are the "true" quality of the purchased good (assumed to equal the econometrician's estimate), which enter the experience utility. Given a consumer's choice of good j, the quantity relevant for welfare is the experience utility of good j. For a given good j in market t, we can compute consumer i's experience utility $\tilde{u}_{ijt}$ with an offset term that depends on the discrepancy between perceived and true qualities and the estimated coefficient on quality.

$$\begin{aligned}\Delta q &:= q_{perceived} - q_{true} \\ \tilde{u}_{ijt} &= u_{ijt} - \beta_1 \Delta q_{ijt}\end{aligned}$$

The welfare for consumer i in market t is then

$$\begin{aligned}W_{it} &= E_\epsilon[u_{ij^*t}] - E_\epsilon[\Delta q_{ij^*t}] \\ &= E_\epsilon[\max_j \{u_{ijt}\}] - E_\epsilon[\Delta q_{ij^*t}] \\ &= \bar{W}_{it} - \sum_{J_t} s_{ijt}(\beta_1 \Delta q_{ijt}),\end{aligned}$$

where j^* is chosen based on perceived quality, and $\bar{W}_{it}$ is the welfare computed under the assumption that consumers care about decision utility.

9.2 Relationship between quality and rating for fake review purchasers

A product j with quality q_j receives organic reviews such that its rating $R_j = q_j$ deterministically. Fake reviews shift ratings such that R_j lies above q_j. The impact of fake reviews on ratings, $R_j - q_j$, is governed by a beta distribution with mean 0.5 that is scaled to lie on the interval $[q_j, 1]$. Formally, $R = q + (1-q)\nu$, where $\nu \sim Beta(\alpha, \beta)$ and $\alpha = \beta$ such that $E[\nu] = 0.5$. Figure 17 describes the shape of the distribution of R_j for a given q_j. Figure 18 depicts the joint distribution of (q_j, R_j).

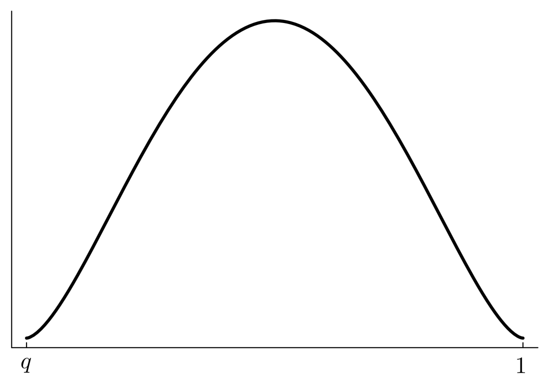


Figure 17: Distribution of R_j with fake reviews.

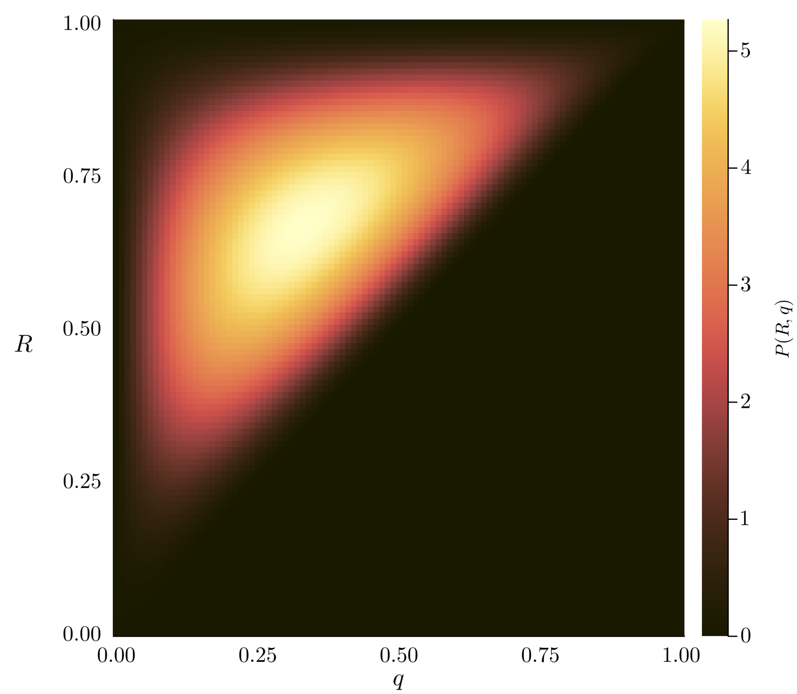


Figure 18: Joint distribution of quality and R

9.3 Posteriors under beta-distributed priors

The consumer's prior beliefs of quality are distributed beta with parameters α_F, β_F for $F \in \{0, 1\}$:

$$\mu_{F,q} = \frac{(p_{F,q})^{\alpha_F - 1}(1 - p_{F,q})^{\beta_F - 1}}{B(\alpha_F, \beta_F)}$$

Given that $F_j = F$ and $q_j = q$, the probability that the first N_j reviews include N_{j0} bad reviews and N_{j1} good reviews is

$$P(N_{j0}, N_{j1}|q_j = q, F_j = F) = \binom{N_{j0} + N_{j1}}{N_{j1}}(P_j^q)^{N_{j1}}(1 - P_j^q)^{N_{j0}}$$

Conditional on $F = F_j$, the consumer's posterior distribution for product j with N_{j0} bad reviews and N_{j1} good reviews is a beta distribution with parameters $N_{j1} + \alpha_{F_j}, N_{j0} + \beta_{F_j}$:

$$\begin{aligned}
P(q_j|N_{j0}, N_{j1}, F_j) = & \; P(q_j|N_{j0}, N_{j1}, F_j) \\
= & \; \frac{P(N_{j0}, N_{j1}|q_j, F_j)\mu_{F_j q_j}}{P(N_{j0}, N_{j1}|F_j)} \\
= & \; \frac{\binom{N_j}{N_{j1}} p_{F_j q_j}^{N_{j1}}(1 - p_{F_j q_j})^{N_{j0}}\mu_{F_j q_j}}{\sum_{q \in \mathcal{Q}} \binom{N_j}{N_{j1}} p_{F_j q}^{N_{j1}}(1 - p_{F_j q})^{N_{j0}}\mu_{F_j q}} \\
\approx & \; \frac{p_{F_j q_j}^{N_{j1}}(1 - p_{F_j q_j})^{N_{j0}}\mu_{F_j q_j}}{\int_{q=0}^1 p_{F_j q}^{N_{j1}}(1 - p_{F_j q})^{N_{j0}}\mu_{F_j q} dq} \\
= & \; \frac{p_{F_j q_j}^{N_{j1}}(1 - p_{F_j q_j})^{N_{j0}} p_{F_j q_j}^{\alpha_{F_j} - 1}(1 - p_{F_j q_j})^{\beta_{F_j} - 1} B(\alpha_{F_j}, \beta_{F_j})^{-1}}{\int_{q=0}^1 p_{F_j q}^{N_{j1}}(1 - p_{F_j q})^{N_{j0}} p_{F_j q}^{\alpha_{F_j} - 1}(1 - p_{F_j q})^{\beta_{F_j} - 1} B(\alpha_{F_j}, \beta_{F_j})^{-1} dq} \\
= & \; \frac{p_{F_j q_j}^{N_{j1} + \alpha_{F_j} - 1}(1 - p_{F_j q_j})^{N_{j0} + \beta_{F_j} - 1}}{\int_{q=0}^1 p_{F_j q}^{N_{j1} + \alpha_{F_j} - 1}(1 - p_{F_j q})^{N_{j0} + \beta_{F_j} - 1} dq} \\
= & \; \frac{p_{F_j q_j}^{N_{j1} + \alpha_{F_j} - 1}(1 - p_{F_j q_j})^{N_{j0} + \beta_{F_j} - 1}}{B(N_{j1} + \alpha_{F_j}, N_{j0} + \beta_{F_j})}.
\end{aligned}$$

The consumer's unconditional posterior distribution is:

$$P(q_j|N_{j0}, N_{j1}) = \sum_{F_j \in \{0,1\}} \frac{p_{F_j q_j}^{N_{j1} + \alpha_{F_j} - 1}(1 - p_{F_j q_j})^{N_{j0} + \beta_{F_j} - 1}}{B(N_{j1} + \alpha_{F_j}, N_{j0} + \beta_{F_j})} P(F_j).$$

Computation for Small Probabilities

Note that $p_{F_jq}^{N_j^+}(1-p_{F_jq})^{N_j^-}\mu_{F_jq}$ tends to be very small, especially when N_j is large. Denote this term by A_{jq} and observe that:

$$\log\left(\sum_{q\in\mathcal{Q}} A_{jq}\right) = \log\left(A_{jq'}\right) + \log\left(\sum_{q\in\mathcal{Q}} \exp\left(\log\left(A_{jq}\right) - \log\left(A_{jq'}\right)\right)\right), \tag{15}$$

$$\tag{16}$$

where q' is a reference quality and $\log(A_{jq})$ is numerically straightforward to compute for any q:

$$\log(A_{jq}) = N_j^+ \log(p_{F_jq}) + N_j^- \log(1-p_{F_jq}) + \log(\mu_{F_jq}). \tag{17}$$

Define $B_{jq} := N_j^+ \log(p_{F_jq}) + N_j^- \log(1-p_{F_jq})$ so that:

$$\begin{aligned}\log\left(\sum_{q\in\mathcal{Q}} A_{jq}\right) &= B_{jq'} + \log(\mu_{F_jq'}) + \log\left(\sum_{q\in\mathcal{Q}} \exp\left(B_{jq} - B_{jq'} + \log(\mu_{F_jq}) - \log(\mu_{F_jq'})\right)\right) \\ &= B_{jq'} + \log\left(\sum_{q\in\mathcal{Q}} \exp\left(B_{jq} - B_{jq'} + \log(\mu_{F_jq})\right),\right)\end{aligned}$$

9.4 Hedonic model of product rank for dynamic counterfactuals

A product listing's rank on Amazon is affected by the sales of the product and its competitors. Accounting for this is important in estimating the full impact of counterfactual policies, as the counterfactual changes in perceived quality affects not just current shares but also future demand through the changes in ranks. To capture this feedback mechanism, we conduct dynamic simulations that estimate the demand in each period using counterfactual product ranks, which are predicted using past-period counterfactual shares. The counterfactual ranks are predicted using estimates from a hedonic model of product ranks based on past shares, past reviews, and current sponsorship status. Table 6 shows the estimates from the hedonic model. Among the lagged variables, the most significant predictors were the market shares and number of good reviews in the past two weeks.

Table 6: Hedonic model of product rank

L1.Log Shares	0.262***	0.281***	0.206***	0.276***
	(18.21)	(18.84)	(15.35)	(18.16)
L2.Log Shares	0.160***	0.177***	0.142***	0.177***
	(11.96)	(12.84)	(10.99)	(12.48)
L1.Log N. Good Reviews	0.100***			
	(14.90)			
L2.Log N. Good Reviews	0.0717***			
	(11.08)			
L1.Cumulative rating		0.0745***		0.0687***
		(9.09)		(8.23)
L2.Cumulative rating		0.0686***		0.0703***
		(8.76)		(8.27)
L1.Weekly rating			0.0232***	0.0200***
			(4.88)	(4.00)
L2.Weekly rating			0.0133**	0.00175
			(2.91)	(0.37)
L1.Log Cumulative N. Reviews		0.105***		0.0900***
		(15.76)		(12.02)
L2.Log Cumulative N. Reviews		0.0758***		0.0595***
		(11.82)		(8.41)
L1.Log Weekly N. Reviews			0.0592***	0.0137
			(8.28)	(1.64)
L2.Log Weekly N. Reviews			0.0304***	0.0221**
			(4.27)	(2.89)
Sponsored	0.476***	0.469***	0.489***	0.471***
	(13.17)	(12.99)	(13.54)	(13.06)
Constant	-1.438***	-1.296***	-1.364***	-1.383***
	(-6.57)	(-5.90)	(-6.19)	(-6.28)
Product FEs	Yes	Yes	Yes	Yes
Observations	317472	317472	317472	317472

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

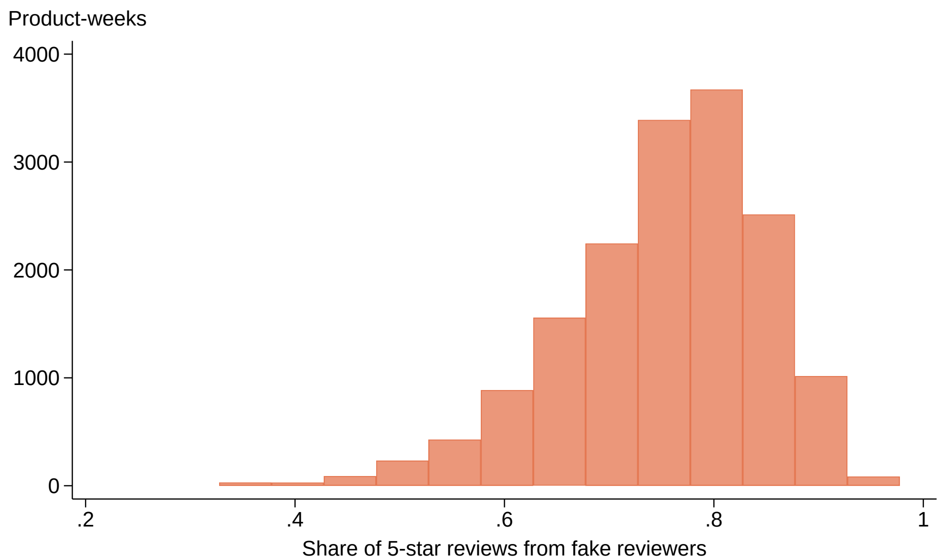


Figure 19: Share of 5-star reviews that come from fake reviewers

Figure 20: Median price elasticity with varying beliefs on prevalence of fake review purchasers