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

Ashvin Gandhi
UCLA & NBER

Brett Hollenbeck
UCLA

Zhijian Li
Northwestern

November, 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.

*We thank Joel Waldfogel, Sherry He, Jessica Fong, Eddie Ning, Jinzhao Du, Ben Vatter, Yeşim Orhun, Andrey Simonov, and Tin Cheuk Leung 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, Columbia GSB, the FTC Microeconomics Conference, the Bass Forms Conference, the IIOC Conference, the Hal White Antitrust Conference, the Summer Institute in Competitive Strategy, the Southern California Strategy Conference, the Quantitative Marketing and Economics Conference, and the BIOMS Conference.

1 Overview

User-generated ratings and reviews are a core feature behind of the success of online marketplaces (Cabral and Hortacsu, 2010; Tadelis, 2016; Einav et al., 2016). These systems solve the asymmetric information problem by allowing sellers to establish reputations. Surveys show that an overwhelming majority of consumers consult reviews before making purchases, both online and offline. As a result, reputation systems have large impacts on marketplace success and seller outcomes, not just online but in many settings such as restaurants, hotels, and healthcare. The importance of these mechanisms creates a powerful incentive for sellers to manipulate their ratings, and recent research has documented that rating manipulation using fake reviews purchased by the seller is widespread (He et al., 2022b; FTC, 2023).

In this paper, we study how fake reviews impact outcomes for sellers, online marketplaces, and ultimately consumers. We propose a framework describing two primary channels by which ratings manipulation can shift outcomes.

The first channel is misinformation. By inflating ratings, fake reviews misinform consumers and may mislead them into make different and possibly worse decisions. The presence of misinformation in markets can also shift equilibrium prices. Products purchasing fake reviews appear higher quality and can increase profits by increasing their prices. This harms even consumers who would have bought these products even absent fake reviews. Increased competition from fake review purchasers could also benefit some consumers by causing honest sellers to lower prices.

The second, and more novel, mechanism is that the widespread presence of fake reviews may cause consumers to generally mistrust ratings. This change in beliefs may impede efficient search by lessening the ability of the ratings system to solve the asymmetric information problem. As a result, consumers may make worse purchase decisions than if they could fully trust product ratings. At the same time, if mistrust in ratings makes high-quality products less able to differentiate themselves from low-quality products, this may benefit consumers through increased price competition.

The relative magnitude of these different forces are unknown, and thus the net impact on aggregate welfare is ambiguous.[1] We seek to quantify these mechanisms empirically in the context of fake product reviews on Amazon.com. In recent years, this context has generated widespread interest by consumer protection regulators around the globe. The FTC, the UK CMA, the European Commission, and others are all investigating the potential consumer harms from rating manipulation and in some cases have proposed laws or other measures in response (FTC, 2019; CMA, 2020). To measure these impacts, we first formalize the mechanisms described above in a model of the equilibrium outcome of rating manipulation. Then, we bring to this model novel data on which Amazon products are using fake reviews and at what times, as well as data from a set of incentivized survey experiments designed to elicit consumer beliefs about the prevalence of fake review activity.

We first model how Bayesian consumers form beliefs about product quality from ratings, taking into account 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 and beliefs about how likely those reviews are to be fake. We use this to show how the misinformation and mistrust effects impact consumers and sellers.

To estimate this model empirically, we first need data on the true prevalence of 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.[2] 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

[1] 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

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

from Amazon on their ratings, reviews, sales ranks, prices, and advertising behavior.

Estimating our model also 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 as 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 removed or prevented 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 that consumers are on net harmed from sellers manipulating their ratings. The net loss in consumer welfare is around 0.64% of the median product purchase price or a loss of about $0.17 per consumer per week. Price competition and consumer beliefs about the trustworthiness of ratings plays a large role in this result. The presence of fake reviews allows the median fake review purchaser to raise prices by $0.33, and the median honest product lowers their prices by $0.07.

This net effect masks important differences in the impacts of the two mechanisms we study. When isolated, the misinformation effect of fake reviews causes a much larger decrease in consumer welfare as consumers are led to buy lower-quality products. By contrast, when fake reviews are not present but consumers mistrust ratings, consumers are actually made better off. When consumers are mistrustful of reviews, sellers react to the increase in competition by decreasing prices. When both effects are present, the benefits of mistrust partially offset the large welfare harms from misinformation and the result is in a fairly modest overall decrease.

Finally, we find that the platform would benefit significantly if it removed all fake reviews and shifted consumer beliefs to reflect this. That is, if they could eliminate both misinformation and mistrust. When isolated, we find that the platform benefits from misinformation, which causes inflated product ratings and demand. By contrast, platform revenue is substantially worse under mistrust as consumers shift purchases offline. Because of this, if the platform were to simply delete fake reviews without consumers adjusting their beliefs, its overall revenue would fall. From the platform's perspective, that is, simply removing fake reviews would 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).) Similarly, Johnen and Ng (2024) considers the welfare gains from sellers lowering their prices to induce positive ratings. 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.[3] 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.[4] 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 A Simple Model of Misinformation and Mistrust

In this section, we illustrate the different ways that rating manipulation can impact consumer choices and firm outcomes. We present a simple model in which consumers make purchases based on observed product features and user ratings that provide a signal of quality. We

[3]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).

[4]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.

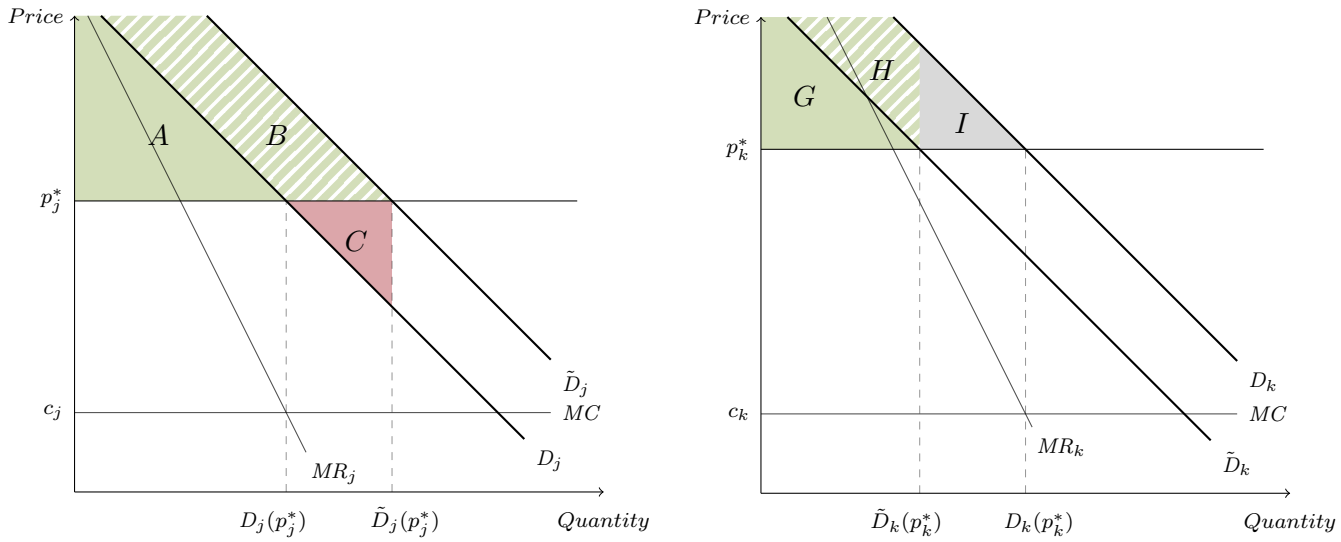
divide our analysis into two distinct effects. The first, which we refer to as the "misinformation effect" of rating manipulation, is that fake reviews provide false information that can mislead consumers into making different purchasing decisions. This is the direct effect that purchasing fake reviews has on a product's sales and the sales of its competitors. The second, which we refer to as the "mistrust effect," is the change in outcomes that results from consumer beliefs that some reviews are fake. Mistrust is a more systemic effect, determined by the overall prevalence of fake review purchasing and not the specific purchasing of any individual product. 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 fake review purchasers and honest products.

2.1 Misinformation

We model consumers' utility from a product j as decreasing in price (p_j) and increasing in 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 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 reflects the true quality. We denote j purchasing or not purchasing fake reviews by F_j and $\neg F_j$, respectively.

Our assumptions imply that in a world without fake reviews, rational consumers will interpret a product's rating to be its quality. We describe a consumer as being "trusting" if they interpret reviews in this way. To best illustrate the effect of misinformation, we first consider how fake reviews impact a market with trusting consumers. Such circumstances

Figure 1: Effect of Misinformation (No Price Changes)



(a) Fake Review Purchaser

(b) Non-Purchaser

might reasonably describe settings in which ratings manipulation is too rare, too new, or too difficult to detect, such that consumers have not yet developed meaningful mistrust.

We consider a market with two competing products, j and k. When qualities are observed by consumers, the demand for product j is $D_j(p_j, q_j, p_k, q_k)$. However, since consumers cannot observe qualities directly, they purchase based on observable ratings. Trusting consumers believe $R_j = q_j$ and $R_k = q_k$, so their demand is characterized by $D_j(p_j, R_j, p_k, R_k)$.

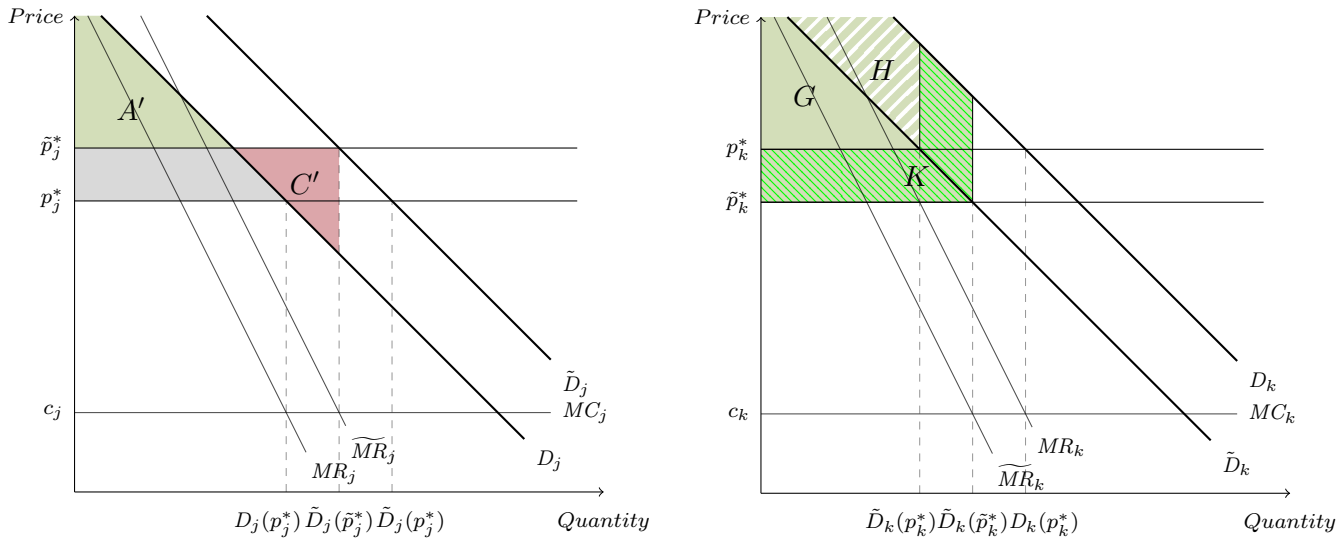
If product j purchases fake reviews, then this increases R_j above q_j and shifts out the demand curve for product j and shifts in the demand curve for competitor product k. Figure 1 shows the effect of these demand shifts when holding prices fixed at the level that would have prevailed without fake reviews. The demand curves D_j and D_k are those that would occur absent fake reviews—i.e., when R_j and R_k accurately reflect q_j and q_k—while $\tilde{D}_j$ and $\tilde{D}_k$ characterize consumer demand given that j purchases fake reviews. Note that while fake reviews cause consumers to purchase according to $\tilde{D}_j$ and $\tilde{D}_k$, the utility actually realized from their purchases are characterized by D_j and D_k. Put simply, the misinformation from fake reviews causes consumers to purchase according to demand curves that do not reflect their informed preferences.

For product j, this entails an increase in quantity demanded from $D_j(p_j^*)$ to $\tilde{D}_j(p_j^*)$, 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. For the $D_j(p_j^*)$ consumers who would have purchased j even absent fake reviews, 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 region C represents forgone utility from making a sub-optimal purchasing decision due to misinformation.

Product k, on the other hand, experiences a reduction in demand from $D(p_k^*)$ to $\tilde{D}_k(p_k^*)$, which reduces profits by $\left(p_k^* - c_k\right)\left(D_k(p_k^*) - \tilde{D}_k(p_k^*)\right)$. Consumers purchasing based on $\tilde{D}_k$ anticipate receiving consumer surplus G. However, these $\tilde{D}_k(p_k^*)$ consumers underestimate their surplus by H because alternative j is actually worse than its ratings suggest. Of course, these consumers would have purchased k even absent fake reviews, so H does not represent a real benefit. In contrast, the $D_k(p_k^*) - \tilde{D}_k(p_k^*)$ consumers induced by fake reviews to purchase j instead of k experience a real harm shown in region I.[5]

[5]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.

Figure 2: Competitive Responses to Misinformation



(a) Fake Review Purchaser

(b) Non-Purchaser

Competitive Responses Of course, both firms should adjust their prices in response to j purchasing fake reviews. Figure 2a depicts these competitive responses. The increase in demand from D_j to $\tilde{D}_j$ raises j's optimal price from p_j^* to $\tilde{p}_j^*$.[6] By raising price, j further increases its profit by $\left(\tilde{p}_j^* - c_j\right) \tilde{D}_j(\tilde{p}_j^*)$ and shrinks consumer surplus from $A - C$ to $A' - C'$.[7] Importantly, this price increase harms the $D_j(p_j^*)$ consumers who would have purchased product j even absent fake reviews. It also exacerbates the harms to the $\tilde{D}_j(\tilde{p}_j^*) - D_j(p_j^*)$ consumers still misled into purchasing j even at the higher price. On the other hand, the $\tilde{D}_j(p_j^*) - \tilde{D}_j(\tilde{p}_j^*)$ consumers dissuaded from purchasing j by the price increase actually benefit from the competitive response.

In contrast, the decrease in demand from D_k to $\tilde{D}_k$ lowers k's optimal price from p_k^* to $\tilde{p}_k^*$. By cutting price, k stems its losses to j and earns a profit of $(\tilde{p}_k^* - c_k)\, \tilde{D}_k(\tilde{p}_k^*) > (p_k^* - c_k)\, \tilde{D}_k(p_k^*)$. This also benefits consumers, who see their surplus increase by region

[6] It is important to note that the competitive responses must solve in equilibrium. As j increases its price, this attenuates the inward shift in k's residual demand curve. Likewise, as k decreases its price, this attenuates the outward shift in j's demand curve. Therefore, when incorporating competitive responses, the equilibrium shifts in demand for j and k are smaller than in Figure 1.

[7] In this example with linear demand and fake reviews shifting only the level of demand, $C' = C$, so the welfare loss is simply $A - A'$.

K. Indeed, $\tilde{D}_k(\tilde{p}_k^*)$ who still purchase k in spite of j's fake reviews now receive a discount that makes them better off than if j had not purchased fake reviews. This shows that misinformation is not unambiguously bad for consumers, as competitive responses benefit those still purchasing honest products. Which effects dominate ultimately depends on the relative sizes of both the price and quality elasticities of demand.

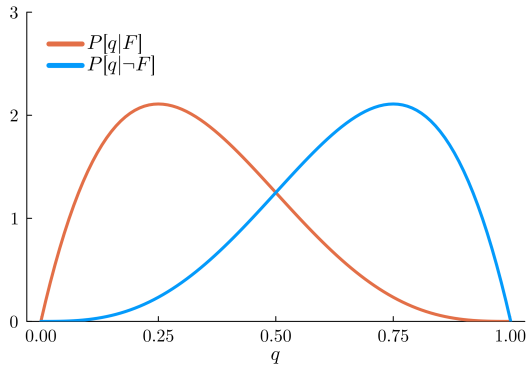
2.2 Mistrust

Thus far, we have modeled consumers as fully trusting reviews in order to isolate the effect of misinformation. However, consumers may become generally aware that some products are manipulating their ratings through fake reviews, even if they cannot precisely identify which products' ratings are inflated. In this section, we explore the implications of precisely this form of mistrust in the rating system. To do this, we model consumers forming expectations product quality that account for the possibility that observed ratings are manipulated to exceed true qualities. Note that we largely suppress product subscripts in order to emphasize that the effect of mistrust works through 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.

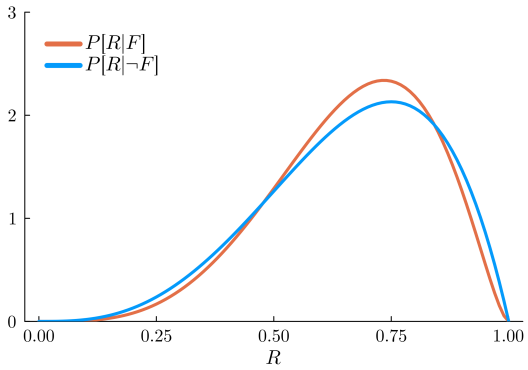
We start by modeling a consumer who cannot identify which products are purchasing fake reviews but has rational expectations about the prevalence of fake reviews. In considering a product with rating R, the mistrustful consumer anticipates some probability $P(F|R) > 0$ that the product purchased fake reviews. If it did, then its rating its rating is inflated, so the expected quality $E[q|R, F]$ is less than R. If it didn't, then R accurately reflects quality. Therefore, the mistrustful consumer forms an expectation about quality that places weight $P(F|R)$ on $E[q|R, F]$ and weight $1 - P(F|R)$ on R:

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

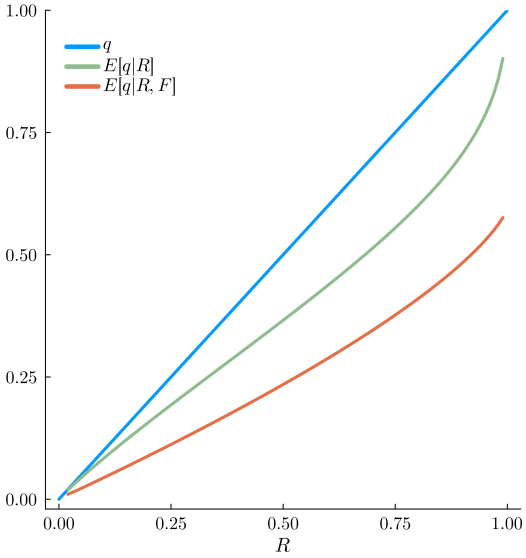
Figure 3: An Illustrative Example



(a) Quality



(b) Ratings



(c) 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 joint distribution of q and R.

Figures 3 provides an illustrative example in which 50% of products purchase fake reviews. In this example, the products that purchase fake reviews tend to have lower qualities (Figure 3a), and in doing so, it improves their ratings to be fairly similar to the ratings for products that don't (Figure 3b). See Appendix 9.2 for details.

Figure 3c illustrates equation (1) characterizing how a Bayesian consumer with rational

expectations infers quality from R. The top line shows R, the quality that the consumer would infer if she were trusting or knew with certainty that the product did not purchase fake reviews. The bottom curve gives $E[q|R, F]$, the expected quality that the Bayesian consumer with rational expectations would infer if she knew for certain that the product purchased fake reviews. Finally, the middle curve gives $E[q|R]$, the quality that the consumer infers 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.)

There are a number of instructive features of Figure 3c. The first is that $E[q|R] \leq R$, so mistrust causes consumers to anticipate lower utility from purchasing any product. This makes any product less attractive, and all else equal, should reduce purchasing. In fact, if $E[q|R]$ were simply a parallel shift downward from R, the only effect of mistrust would be to shift demand to the outside good. However, the shift downward is not parallel because the mistrusting Bayesian discounts their expectation differently depending on the product's observed rating. Specifically, the Bayesian consumer discounts their expectations most heavily when a product's rating indicates that it likely purchased fake reviews—i.e., $P(F|R)$ is large—or that the products achieving such a rating through manipulation are particularly bad—i.e., $E[q|R, F]$ is much lower than R.

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 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.

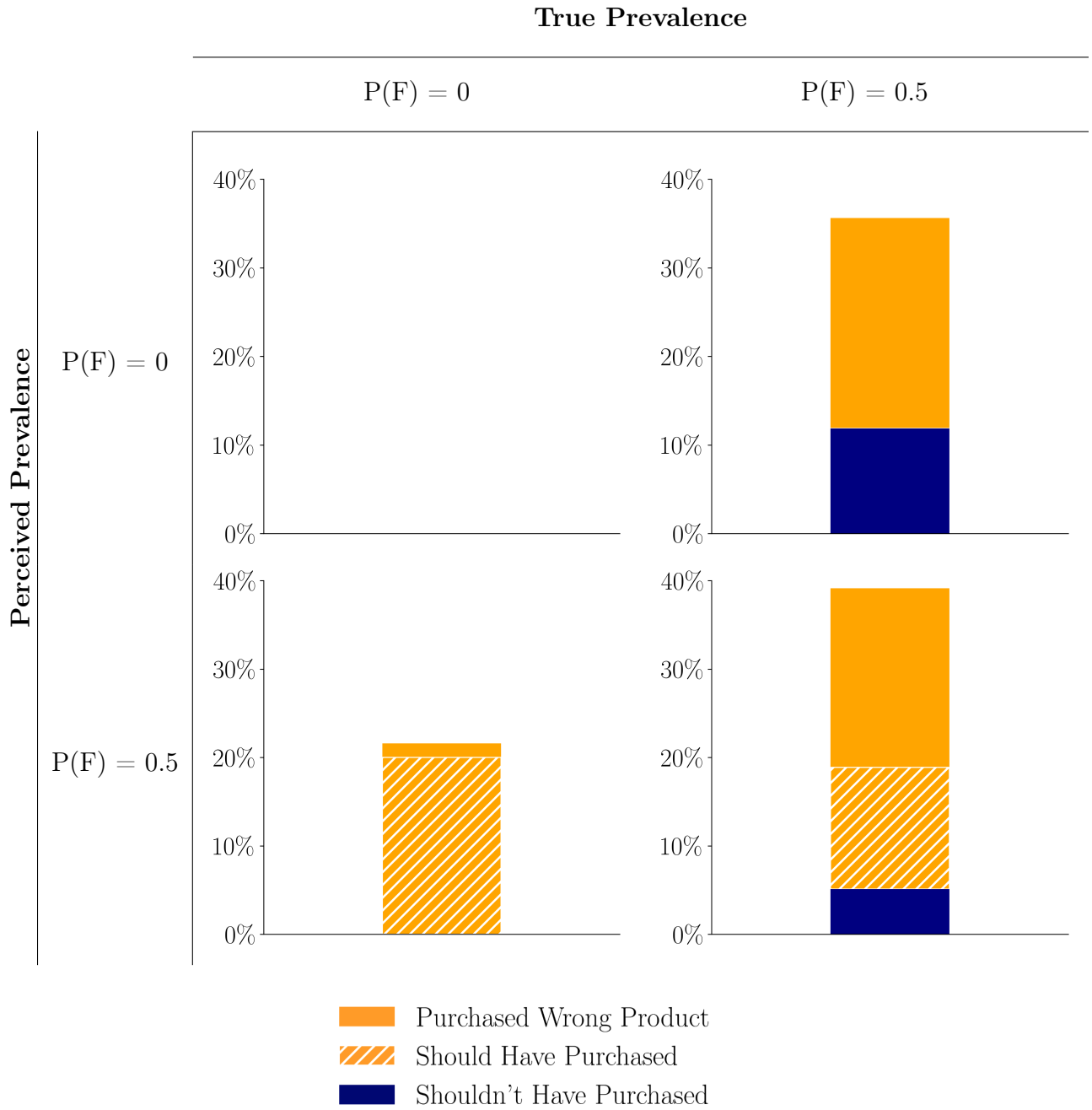
Relaxing Rational Expectations There are a number of reasons that consumers' beliefs about fake reviews may not satisfy rational expectations. For example, consumers may under- or overestimate the prevalence of fake reviews yielding inaccurate beliefs about $P(F|R)$. Likewise, consumers may misunderstand how much fake reviews move R and therefore infer $E[q|R, F]$ incorrectly. Relaxing rational expectations simply requires specifying how the beliefs in equation 1 are determined. In our empirical exercise, we characterize these beliefs using a survey experiment.

Comparing the Effects of Misinformation and Mistrust Of course, both misinformation and mistrust are likely to be present in many markets. Therefore, we return to the illustrative example from Section 2.2 and compare how misinformation and mistrust shift consumer choices. Figure 4 depicts four scenarios in four quadrants, which vary based on the true prevalence of fake reviews (i.e., misinformation) and the perceived prevalence (i.e., mistrust). In the upper-left quadrant, neither misinformation nor mistrust are present, while in the bottom right quadrant, both are present.

When there is only misinformation (upper-right), consumers buy too many product that purchased fake reviews. If fully informed, these consumers would have preferred to purchase other honest products (orange) or not to have purchased at all (blue). When there is only mistrust (bottom-left), the primary distortion in choices is that consumers buy too few products from the marketplace and shift those purchases to the outside option.[8] Finally, when there is both misinformation and mistrust, consumers make all three types sub-optimal choices: they purchase the wrong product, purchase when they should not have, and do not purchase when they should have.

In sum, even this toy example suggests that the ultimate implications of mistrust and misinformation for substitution patterns are highly dependent on many empirical factors, including the shape of consumer demand, the prevalence and magnitude of fake reviews,

[8]Note that neither of the off-diagonal outcomes is a full equilibrium outcome because beliefs and the underlying state of the world are misaligned. These should be interpreted as comparative statics meant to isolate the different mechanisms.



Note: All plots are simulated with 10000 random draws from the Beta distributions and 10000 customers, assuming the outside option quality is 0.5. The randomness from the customers is modeled by $Gumbel(0, 0.1)$.

Figure 4: Percentage of Wrong Choices Under Misinfo and Mistrust

and the distribution of quality for both fake review purchasers and honest products. This complexity underscores the importance of the empirical exercise that we explore in the remainder of our paper.

3 Data

Our aim is to measure the effects of rating manipulation on Amazon on consumer demand and understand how that demand would change under alternative regulatory scenarios. To do this, we require data from Amazon on product characteristics, ratings, sales, and how these vary over time, as well as 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.

Figure 5 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 6 shows the distribution of the products' average ratings, separately based on whether the product is an FRP or an "honest product" (HP). Most products have average ratings between 4 and 5 stars, with the focal products'

Table 1: **Top Categories of Fake Review Purchasers**

Category	Product-weeks
Beauty & Personal Care	106
Health & Household	95
Home & Kitchen	75
Kitchen & Dining	59
Tools & Home Improvement	59
Cell Phones & Accessories	43
Pet Supplies	38
Sports & Outdoors	35
Patio, Lawn & Garden	32
Electronics	27

Figure 5: Distribution of Product Prices

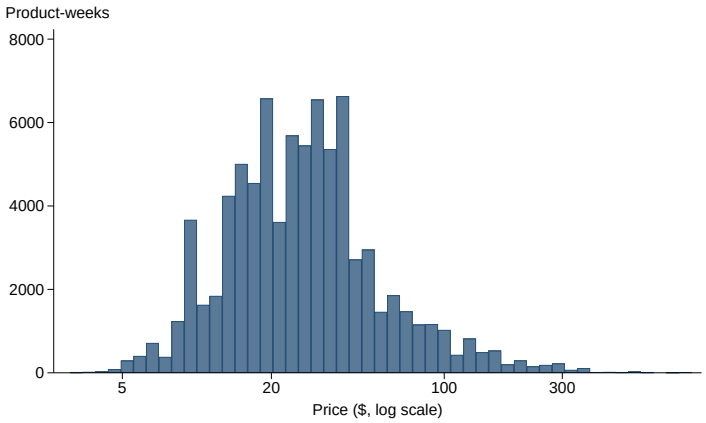
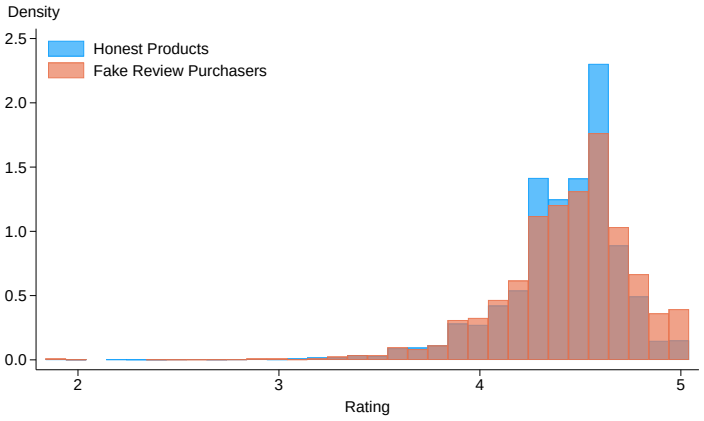


Figure 6: Average rating



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 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 a 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

Table 2: Characteristics of Fake Review Purchasers and Comparison Products

	Count	Mean	SD	25%	50%	75%
Displayed Rating						
Fake Review Purchasers	1346	4.34	0.43	4.13	4.40	4.63
Close Competitors	3153	4.31	0.37	4.15	4.38	4.56
All Products	221923	4.25	0.61	4.01	4.37	4.62
Number of Reviews						
Fake Review Purchasers	1341	193	398	29	77	194
Close Competitors	3153	1266	6122	85	274	905
All Products	221925	378	2019	11	51	217
Price						
Fake Review Purchasers	1354	33.25	45.25	15.46	23.53	35.40
Close Competitors	3153	38.03	47.35	15.94	24.99	39.95
All Products	245415	43.48	190.57	12.99	20.99	38.75
Sponsored						
Fake Review Purchasers	1354	0.17	0.20	0.01	0.09	0.28
Close Competitors	3153	0.33	0.23	0.14	0.30	0.49
All Products	245452	0.09	0.19	0.00	0.00	0.07
Keyword Position						
Fake Review Purchasers	1352	98	62	49	90	136
Close Competitors	3153	97	53	56	92	129
All Products	244160	187	76	133	190	243
Age (months)						
Fake Review Purchasers	1239	9.48	8.26	4.77	7.10	11.09
Close Competitors	3153	21.04	23.40	7.24	12.71	26.19
All Products	245936	22.47	26.15	6.00	12.91	29.61
Sales Rank						
Fake Review Purchasers	1354	174911	307317	32675	95547	202742
Close Competitors	3153	115993	215791	12737	49420	134665
All Products	246051	365923	691609	51652	166437	411728

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).

3.1 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.

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. We examine all the reviews of these products and use a subsample of reviewers to identify reviewers observed leaving multiple five-star reviews for products classified as purchasing fake reviews. We label these reviewers as "fake reviewers". Then, for each product j, we use this to estimate the fraction of j's total five-star reviews that came from fake reviewers. This provides an estimate of the proportion of fake reviews for that product. For the products we observe buying fake reviews, the average estimated share of fake reviews is 56% with a median share of 59%. We provide more details on this procedure in Appendix 9.5.

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 honest products, 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 $P(q \mid N, F) = P(q \mid F)$.[9]

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 3.1 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 excluding 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.

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:[10]

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

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.[11] The average quality of a product that purchases fake reviews is 0.41, 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

[9]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).

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

[11]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.

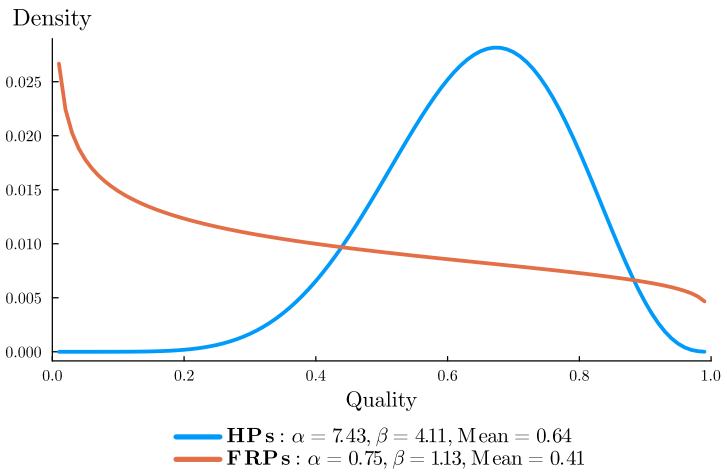


Figure 7: Estimated Priors

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 on only its organic ratings:[12]

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

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

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

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). 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

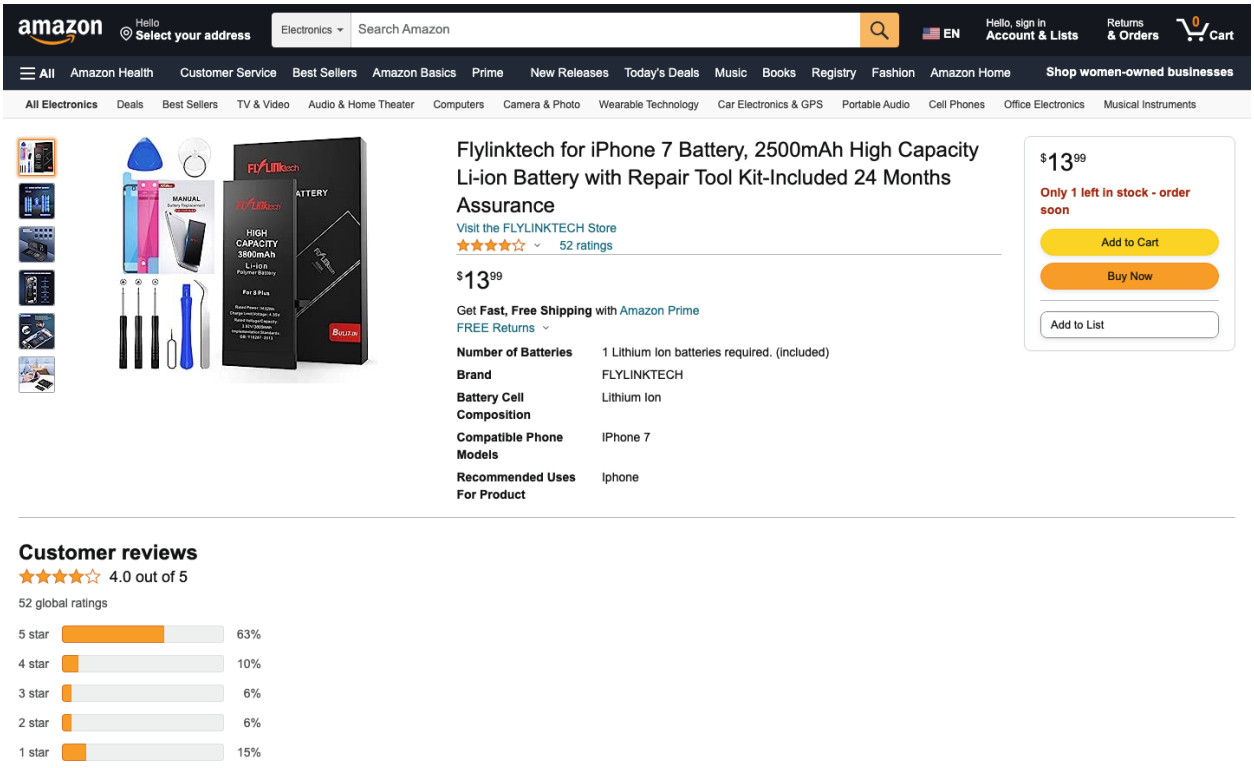


Figure 8: Example product page shown as in survey

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 from each category) are randomly drawn from a set of 1541 fake review purchasing products.

Figure 9: Slider after the respondent selects a probability.

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 Figure 10.

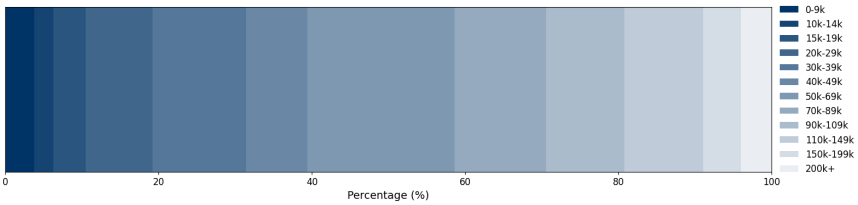
For the initial question, asking "Out of 100 randomly chosen products on Amazon.com, how many would you expect to have purchased fake reviews?" the mean response is 31% and the median is 26%. This is slightly higher than the 20% of products we observe in our data. For the prediction task questions, beliefs 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 11, 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.

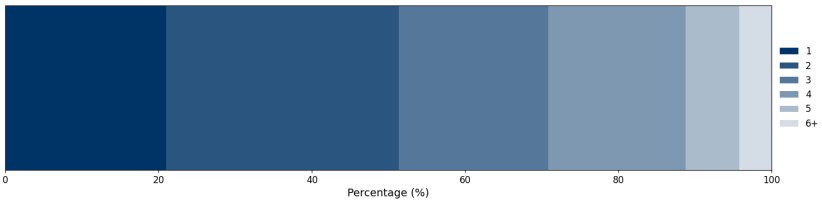
Next, we examine how the mean response varies across the full set of interactions between the average rating and the number of reviews. Figure 12 shows the full results in heatmap form. We see that consumers are especially suspicious of products with very few reviews but a very high average rating. Products 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

Figure 10: **Demographics of Survey Respondents**

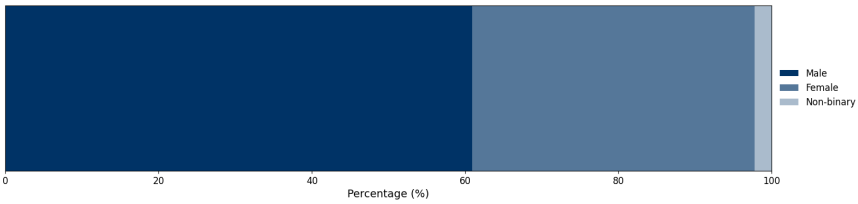
(a) Income



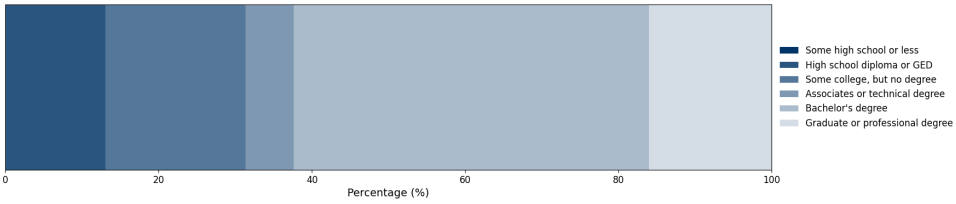
(b) Household Size



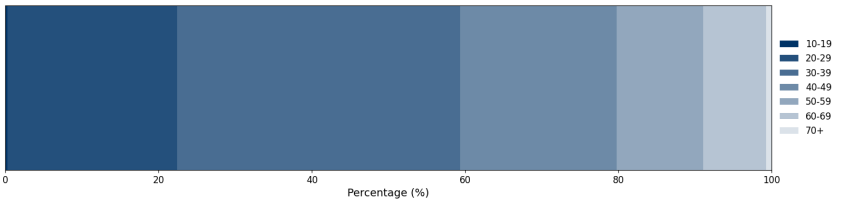
(c) Gender



(d) Education

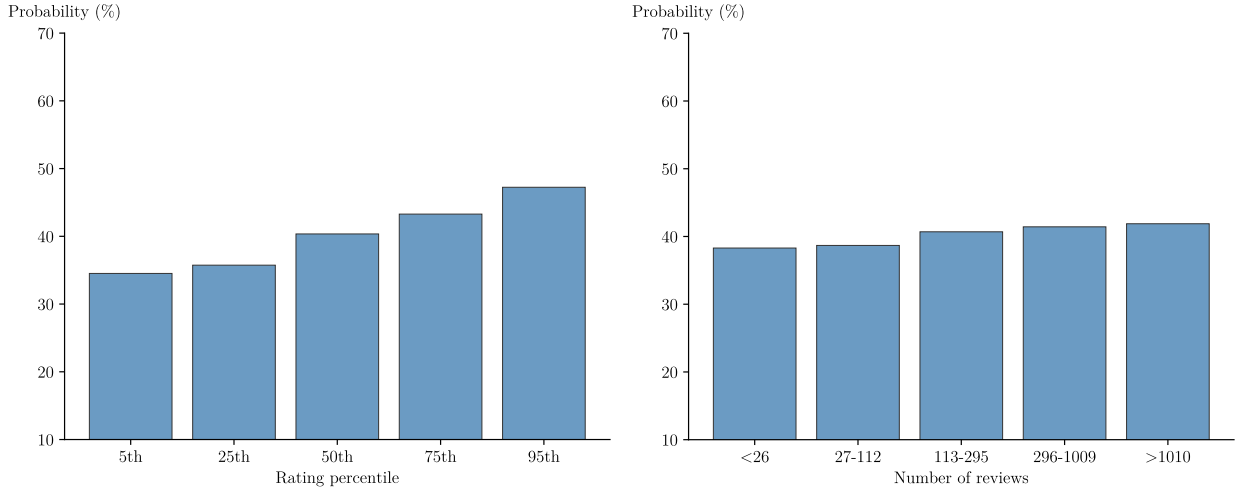


(e) Age



Notes: For subfigure (10d), 0.5% percent of participants put "Prefer Not to Say".

Figure 11: Beliefs About Fake Reviews by Product Characteristics



(a) By Rating

(b) By # Reviews

review likelihood.

We next investigate how the shape of the rating histogram affects the perceived likelihood of fake review purchasing. Respondents gave a higher probability response when shown ratings histograms with higher variances in ratings than with lower variances. However, this pattern is driven by the product pages with very few reviews or low ratings. Details on the histogram variation and results are reported in appendix 10.1.

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, 0% of the respondents correctly responded 0%, and the mean response is 11%. Figure 13 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.

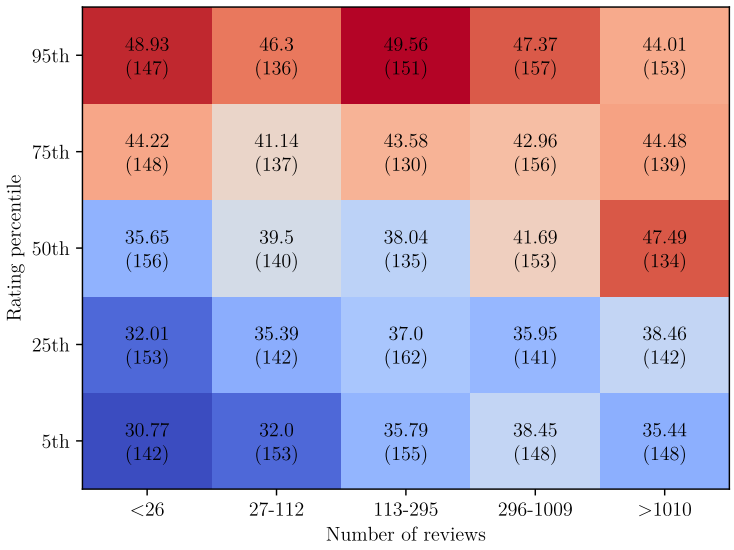


Figure 12: Beliefs by Product Characterstics

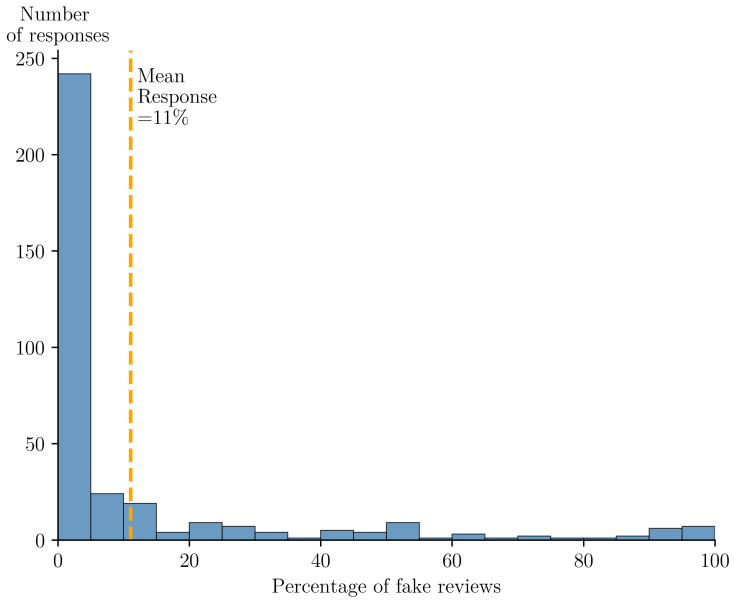


Figure 13: Responses for Amazon Gift Card

5.2 Supplemental Survey: Review Text

To assess the effect of the content of reviews in consumer beliefs, we designed a supplemental survey that is structured similarly to the first survey, but with an additional option for respondents to view a sample of the product's reviews in addition to the product page

during each of the prediction tasks. [13]

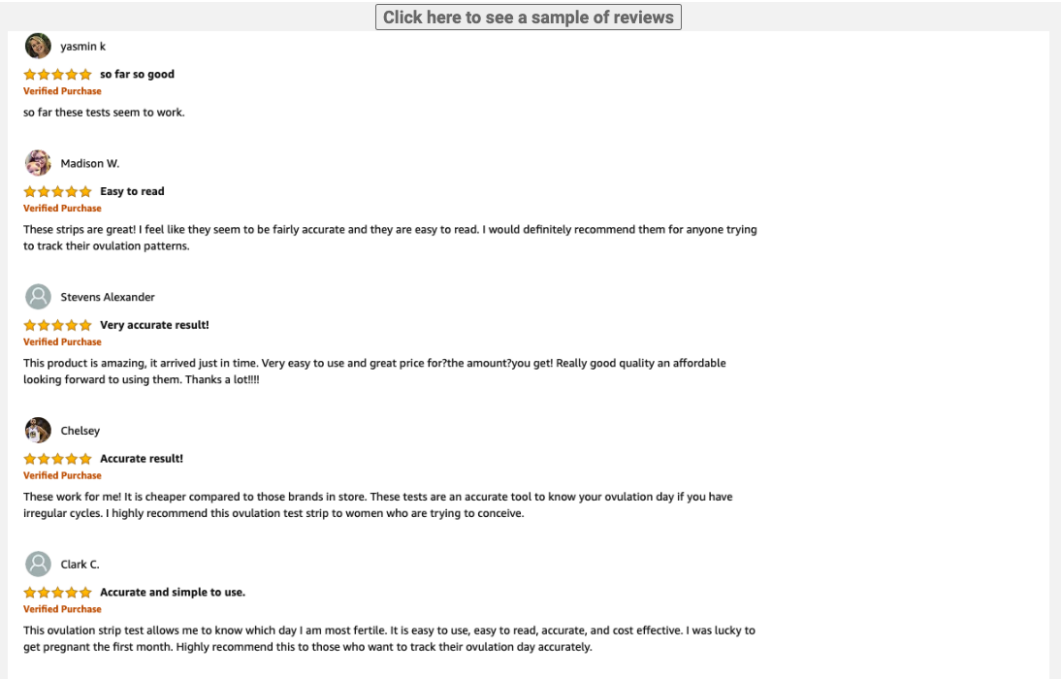


Figure 14: Example of the review page displayed after the respondent chooses to see reviews.

In April 2024, we ran this survey on a different set of 100 participants from the same participant pool on Prolific. Overall, participants clicked to view the review page 86% of the time[14]. We found that when participants clicked to see reviews, the prediction of $P(F_j|N)$ was still largely unaffected by whether the product purchased fake reviews: the mean prediction was 35.2% if participants saw the reviews for a fake review purchaser and 34.8% if they saw the reviews for an honest product. In contrast, if the participant did not click to see reviews, the mean prediction was 48.5% for fake review purchasers and 39.8% for honest products. Our interpretation of these results is that the majority of consumers are unable to identify fake review purchasers with or without looking at review content. However, there may be a small proportion of sophisticated consumers who can spot some features that serve as (weak) signals of fake review purchasing activity, but these signals

[13]For fake review purchasers, we show the first 10 reviews received after date of fake review purchase, which were between December 2019 and June 2020. For the honest products, we select the earliest 10 reviews among our data, which were scraped between August and December 2020.

[14]Click-through rates were almost identical when shown fake review purchasers compared to honest products.

likely do not appear in the review content.

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 price and expected quality as $\binom{\alpha_i}{\beta_{0i}} \sim log\mathcal{N}(\mu, \Sigma)$. The use of a lognormal distribution of individual heterogeneity restricts preferences such that all consumers place positive weight on expected quality and negative weight on price.

We define market at the keyword-week level and denote the set of products in the market as $\mathcal{J}$. To construct this set of competitors, we use our data from several months of scraping keyword search results and calculate the frequency with which products co-occur on the same page of search results. Then, for each focal product we choose the set of up to ten products that co-occur most frequently.

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 4, 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 $\theta = (\alpha, \beta, \Sigma, \rho)$. We include an additional set of supply-side moments using a simple supply-side model consisting of only an intercept term. We also implements a covariance restriction between demand-side and supply-side structural error terms as suggested by Mackay and Miller (2024). Formally the GMM estimator is formed from the combination of these 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. We combine the three sets of moments into $\omega(\theta)$ to construct the GMM Objective function. For all specifications, we employ the second-stage heteroskedasticity robust optimal weighting matrix and the Chamberlain (1987) approximation to the optimal instruments as described in (Conlon and Gortmaker, 2020).

In order to obtain a first-stage estimate to construct the weighting matrix and approximation to the optimal instruments, we need to choose initial instruments. For the simple supply specification we use only the product-level intercept. For demand, we follow a standard approach and use Gandhi & Houde-style instruments constructed from the product characteristics of competing products. We rely on product fixed effects to absorb mean product quality. Thus, we treat 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 3 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 1.5. This is not directly comparable to previous estimates of the elasticity with respect to ratings. We find a mean price elasticity of -1.9 with a median of -1.5. This suggests somewhat inelastic demand, consistent with prior estimates of Amazon product demand (Reimers and Waldfogel, 2017, 2021). We find a negative coefficient on the product age and a negative coefficient on the listing rank, consistent with our expectation of higher demand for newer and higher-ranked products.

Table 3: Results of Demand Estimation

Age	-0.036
	(0.018)
Listing Rank	-0.029
	(0.00096)
μ_{-p}	-2.8
	(0.065)
σ_{-p}	0.16
	(0.011)
μ_q	0.76
	(0.032)
σ_q	0.023
	(0.0074)
Product FEs	Yes
Week FEs	Yes
Gandhi-Houde IVs	Yes
Median Own-Price Elast.	-1.4
Mean Own-Price Elast.	-1.9
Median Own-Quality Elast.	1.5
Mean Own-Quality Elast.	1.5
Observations	81,364

Notes: The random coefficients are parameterized as $\binom{\alpha_i}{\beta_{0i}} \sim log\mathcal{N}(\mu, \Sigma)$ where $\mu = \binom{\mu_{-p}}{\mu_q}$ and $\Sigma = \begin{pmatrix} \sigma_{-p} & 0 \\ 0 & \sigma_q \end{pmatrix}$.

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 3.1 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 15: Average rating, fake review purchasers

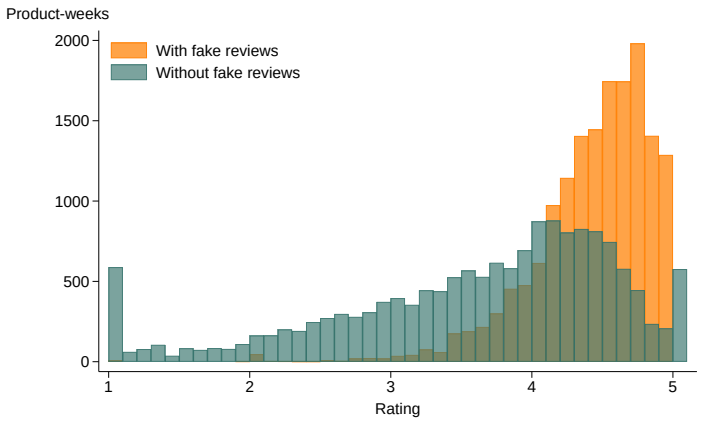
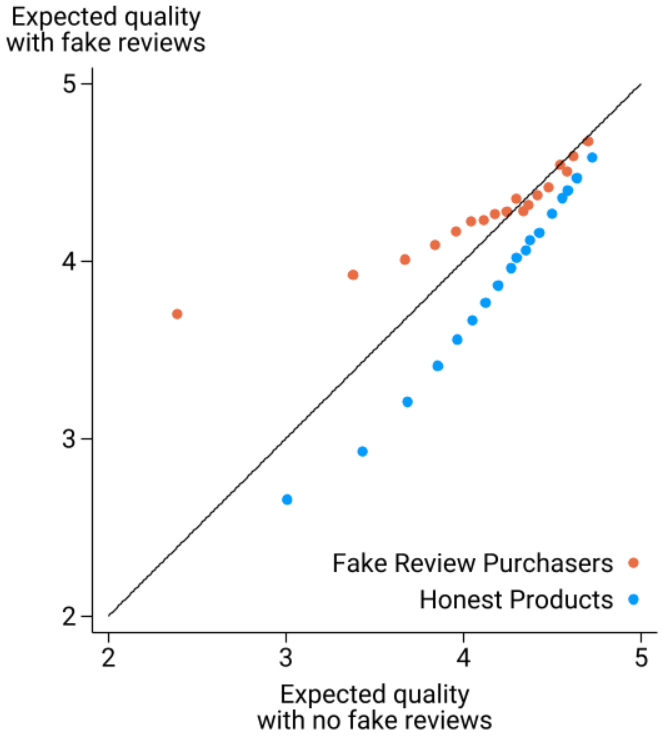


Figure 15 shows the distribution of average ratings for focal products when the fake reviews are included compared to when they are absent. Next, these average product ratings and numbers of reviews are used to compute expected quality. In Figure 16 the perceived qualities without fake reviews are plotted 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 honest products, the presence of fake reviews in consumer beliefs causes their perceived quality to fall as a result of consumer

mistrust.

Figure 16: Perceived Product Quality with and without Fake reviews



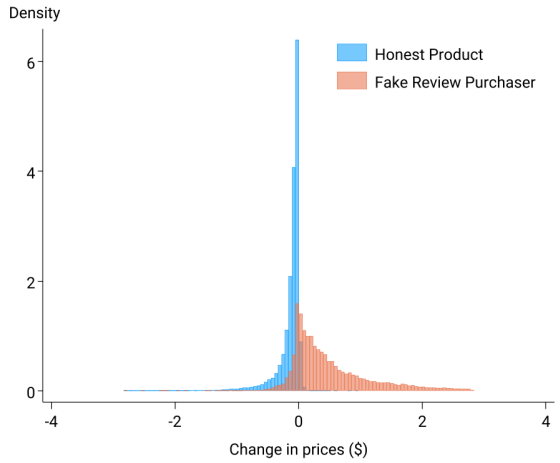
Next, we solve for product demand under the factual and counterfactual set of perceived qualities and allow sellers to adjust their prices. We run the simulated outcomes for all weeks in our data sequentially and note that there is a potentially important feedback loop embedded in the data. If in period 1, an FRP loses sales in the counterfactual relative to an HP, 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 $t+2$ positions. Details and results of the hedonic model estimation are reported in Appendix 10.2.

The full set of results are summarized in Table 4, 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 17 visually depicts the changes in equilibrium prices with fake reviews present vs absent. The median change in prices is an increase of \$0.33 for fake review purchasers and a decrease of \$0.07 for honest products. Fake review purchasers

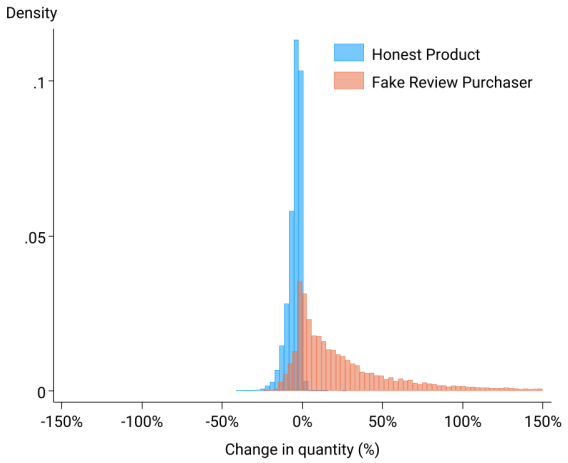
are able to charge higher prices because of the upward adjustment in their ratings, and because of consumer mistrust due to fake reviews, the other products decrease their prices. The net sales-weighted average price difference due to the presence of fake reviews is -$0.05 and there is almost zero net change in sales 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 18. 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.

Figure 17: Counterfactual differences in prices and quantities



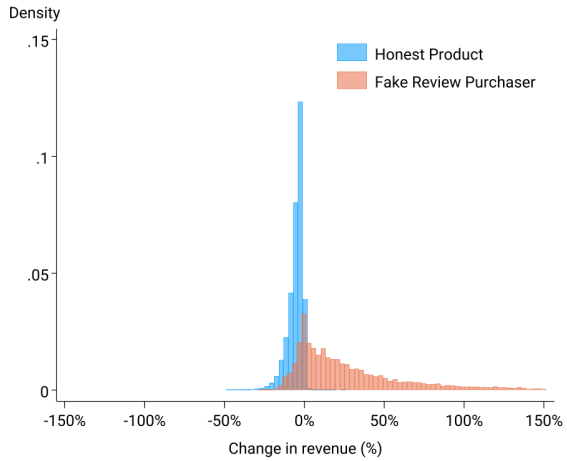
(a) Change in prices. The median change is -$0.07 for Honest Products and $0.33 for Fake Review Purchasers.



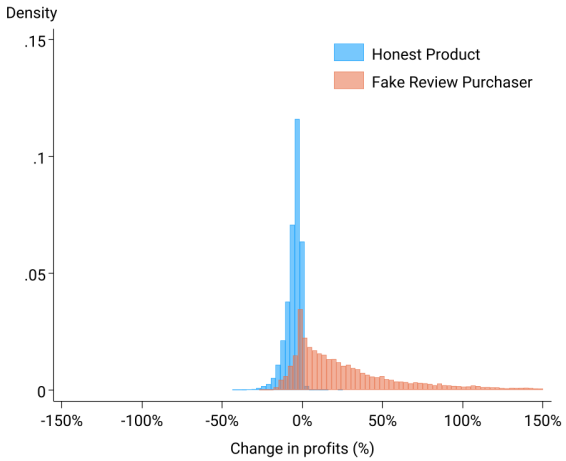
(b) Change in quantities. The median change is -3.7% for Honest Products and 17% for Fake Review Purchasers.

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 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,

Figure 18: Counterfactual differences in revenues and profits



(a) Change in revenues. The median change is -4.1% for Honest Products and 21% for Fake Review Purchasers.



(b) Change in profits. The median change is -4.2% for Honest Products and 21% for Fake Review Purchasers.

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:

$$\begin{aligned}\Delta q &:= q_{perceived} - q_{true} \\ &= \mathbb{E}\left[q|N^+, N\right] - \mathbb{E}\left[q|N^+, N, F, \theta\right] \\ \tilde{u}_{ijt} &= u_{ijt} - \beta_1 \Delta q_{ijt}\end{aligned}$$

The welfare for consumer i in market t is then

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

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 0.64% of the median product purchase price or a loss of about $0.17 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 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 4. We find that misinformation alone causes only a small decrease in consumer welfare. In this scenario, consumers shift their purchases towards Fake Review Purchasers, who charge higher prices and have lower true qualities. However, this negative welfare effect is offset by the large fraction of consumers who still buy Honest Products, whose prices have fallen. When we compare this to the effects of mistrust alone, we find that mistrust slightly increases consumer welfare. Mistrust causes consumers to buy fewer of each type of products but this effect is offset by the gains to consumers from the fact that prices fall via increased competition. On net, when both effects are present,

consumers are harmed and it is clear that a substantial majority of this harm results from misinformation distorting ratings and thus choices. 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 Honest Products offsets the increase in prices from Fake Review Purchasers 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 misinformation 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, simply removing fake reviews could backfire in the short run if consumers are not informed about this or do not find it credible.

Table 4: Outcomes in Counterfactuals

	No FR	Misinfo	Mistrust	Misinfo+Mistrust	
		Floating prices	Floating prices	Fixed prices	Floating prices
Welfare ($)	157,165,661	155,903,258	157,219,372	155,803,143	155,994,229
Platform revenue ($)	18,343,680	18,504,938	18,246,386	18,362,822	18,415,361
FRP average prices ($)	30.69	31.58	30.65	30.69	31.55
HP average prices ($)	37.91	37.78	37.89	37.91	37.76
FRP sales (units)	637,707	847,726	630,224	889,932	842,323
HP sales (units)	4,379,192	4,249,945	4,359,467	4,187,746	4,229,189
FRP profits ($)	12,962,333	17,621,840	12,792,734	17,533,463	17,495,151
HP profits ($)	84,934,324	81,794,001	84,513,993	81,075,406	81,361,446

To further break down the welfare harms of misinformation and mistrust, we decompose the types of choices consumers make when they make suboptimal choices under misinformation and/or mistrust. There are three mutually exclusive ways to make a suboptimal choice: purchasing the wrong product when the optimal choice is another product, not purchasing any product when the optimal choice is a purchase, and purchasing a product when

the optimal choice is not to purchase. For each type of suboptimal choice, we quantify the welfare harms incurred by consumers making such choices. Figure 19 depicts the fraction of consumers who make each type of wrong choice under each counterfactual scenario. The right panels illustrate how the presence of misinformation artificially boosts perceived quality and results in purchases by consumers who should not have purchased, and the bottom panels show that mistrust has the opposite effect by artificially diminishing perceived quality. Misinformation results in many consumers purchasing the wrong product as it inflates the perceived quality of Fake Review Purchasers only, but even mistrust can result in wrong product selections as consumers may be more or less suspicious of different products depending on their rating and number of reviews. Figure 20 quantifies the magnitude of welfare loss arising from each type of mistake, indicating that the mistakes under misinformation are the most harmful to welfare. Comparing the right panels in figures 19 and 20, we find that, with misinformation, the presence of mistrust tend to prevent the mistakes that are the most detrimental to welfare.

7.3 Costs and Benefits to Sellers

Next, we analyze the costs and benefits of fake review purchasing from the sellers' perspective.

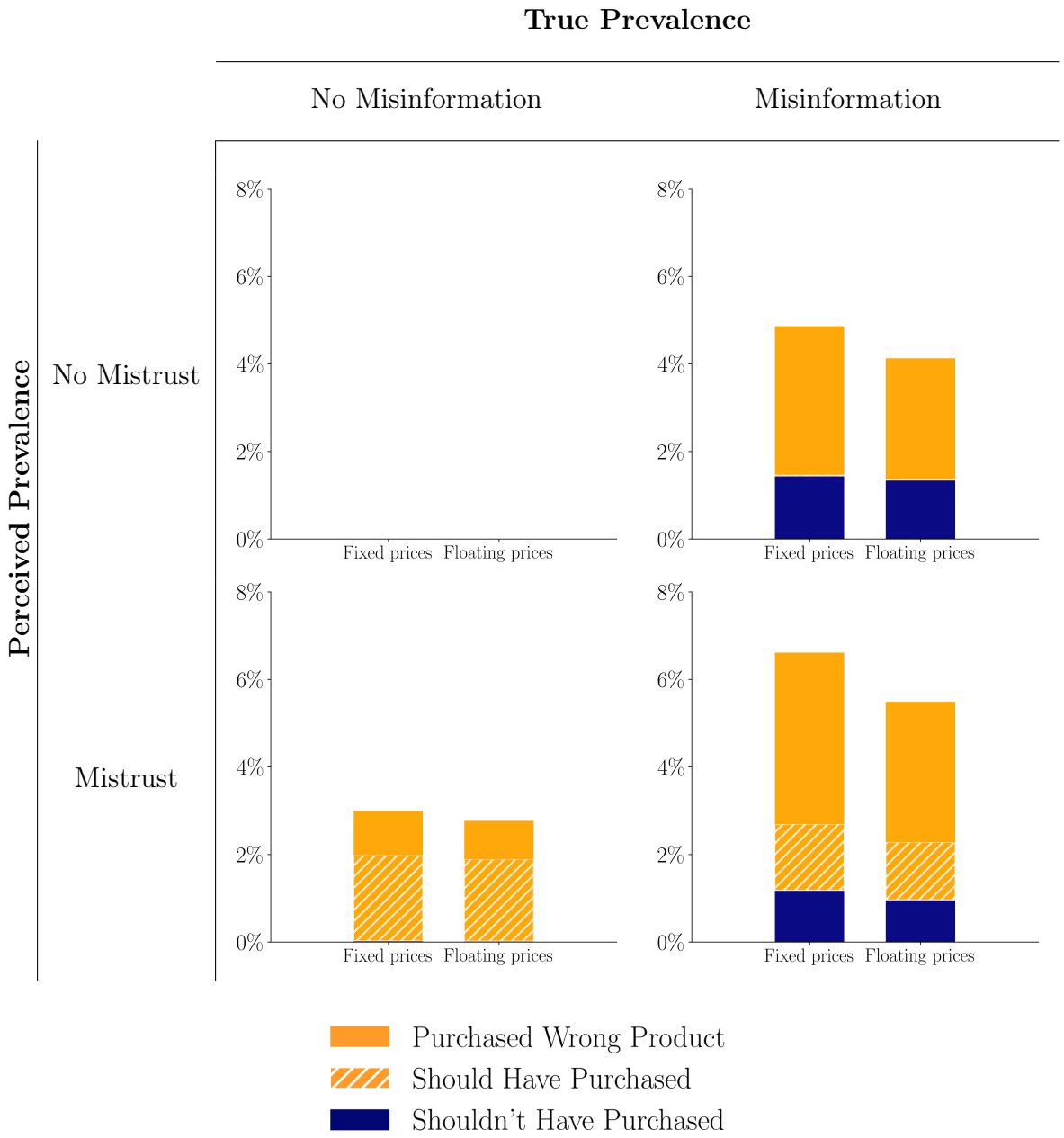
Following He et al. (2022b), we compute the cost of purchasing one fake review as

$$Cost = MC + P(1 + \tau + F_{PP}) + Commission - P(1 - c), \tag{15}$$

where MC is the product's production cost, P is the product's list price, τ is the sales tax rate, F_{PP} is the PayPal fee, $Commission$ is any additional payment from the seller to the reviewer, and c is Amazon's commission. We assume $\tau = 0.0656$ (the average of state and local sales taxes), $F_{PP} = 2.9\%$, $Commission = 0$, $c = 15\%$.[15]

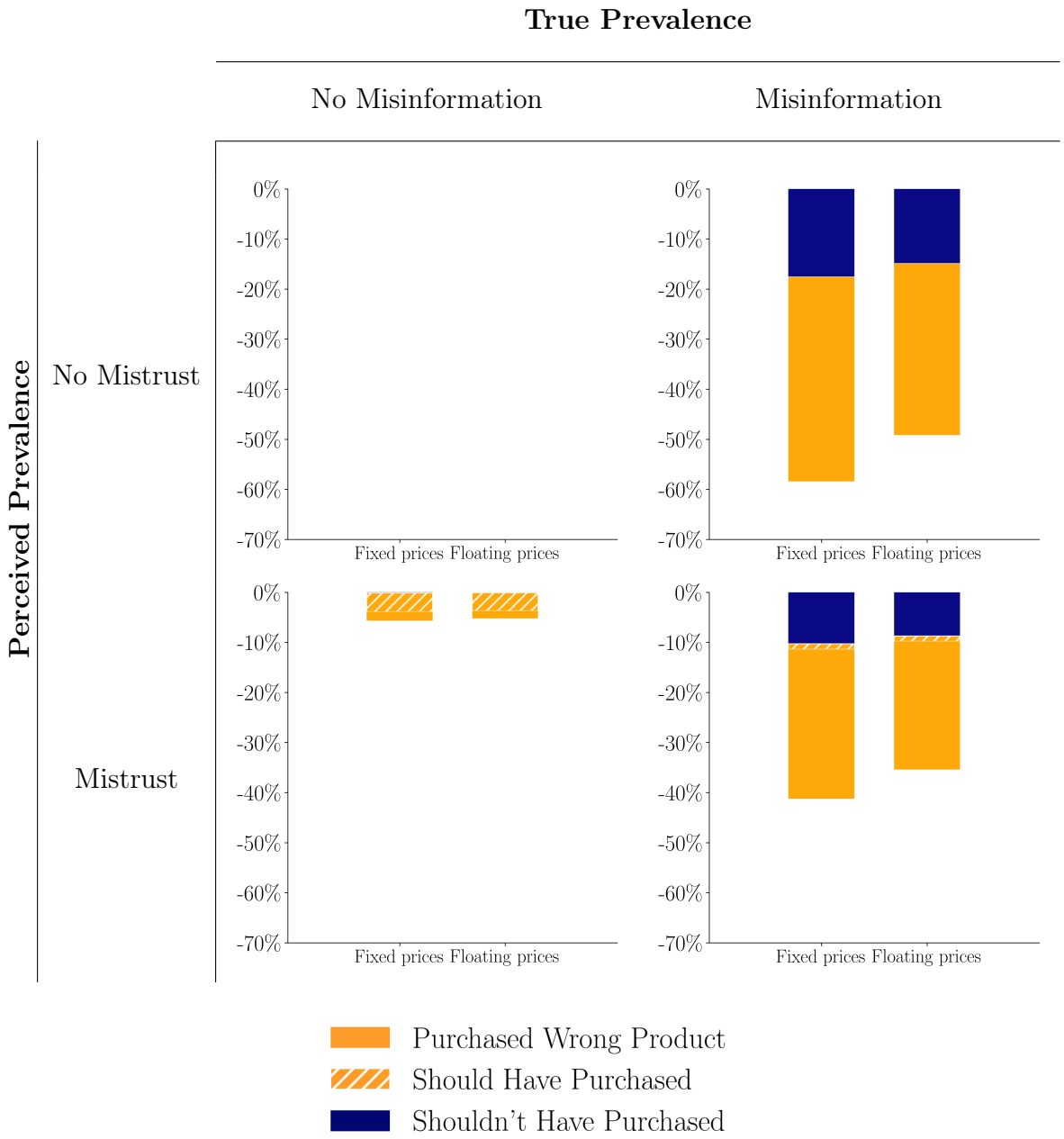
We compute expected qualities before and after a fake review purchase and simulate

[15]The commission is often zero but sometimes \$5 to \$10.



Note: Figure tabulates the number of consumers who make each type of mistakes made under combinations of misinformation and mistrust. In each panel, we first fix prices at the levels with neither misinformation nor mistrust, then allow for re-pricing in equilibrium.

Figure 19: Percentage of Wrong Choices Under Misinformation and Mistrust



Note: Figure tabulates the proportion of welfare harms done by each type of mistake made under each counterfactual scenario. The denominator is the total welfare of all consumers who made mistakes in the given counterfactual scenario, had they chosen correctly.

Figure 20: Welfare Harms Under Misinformation and Mistrust

market outcomes to compute the benefit to sellers of a single fake review. We focus our analysis on the first fake review a seller purchases and, for Fake Review Purchasers, one additional fake review.

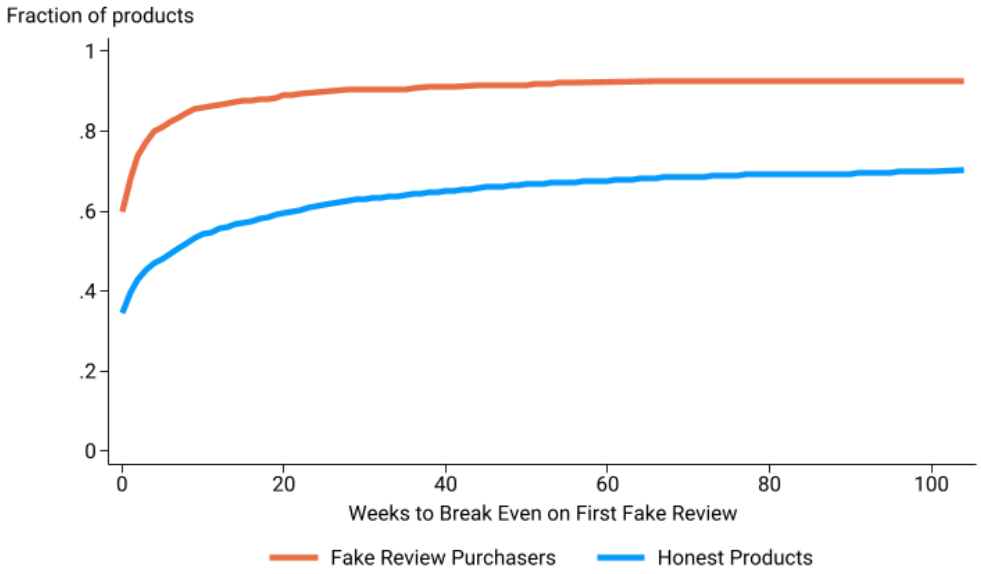
We find that fake review purchase is generally profitable for sellers. Figure 21 reports the cumulative distribution of the number of weeks for products to break even on a fake review purchase. Purchasing a first fake review pays for itself within a year for over 80% of Fake Review Purchasers, and an additional fake review pays for itself for over 70% of Fake Review Purchasers. A would-be first fake review is much less profitable for non-purchasers, who generally have better and more numerous organic ratings.

Next, we compare the benefits from purchasing a fake review to that of purchasing a fake negative review for one's competitor. For each Fake Review Purchaser, we find its closest competitor as determined by the elasticity of demand with respect to the competitor's expected quality, and simulate the market equilibrium under the counterfactual competitor quality. We compare the additional profit to the Fake Review Purchaser to the cost of a negative fake review, which is now simply $P(1 + \tau + F_{PP}) + Commission$. For the median Fake Review Purchaser, the cost of a negative fake review is \$26.48 compared to the cost of a positive fake review of \$11.70, whereas the 4-week profit of a negative fake review is \$5.21 compared to \$55.15 for a positive fake review. Figure 22 compares the profitability of such a fake review to the standard own-product fake review. For most products, it takes much longer for the negative fake review to break even, which indicates that the relative lack of fake negative reviews is economically rational.

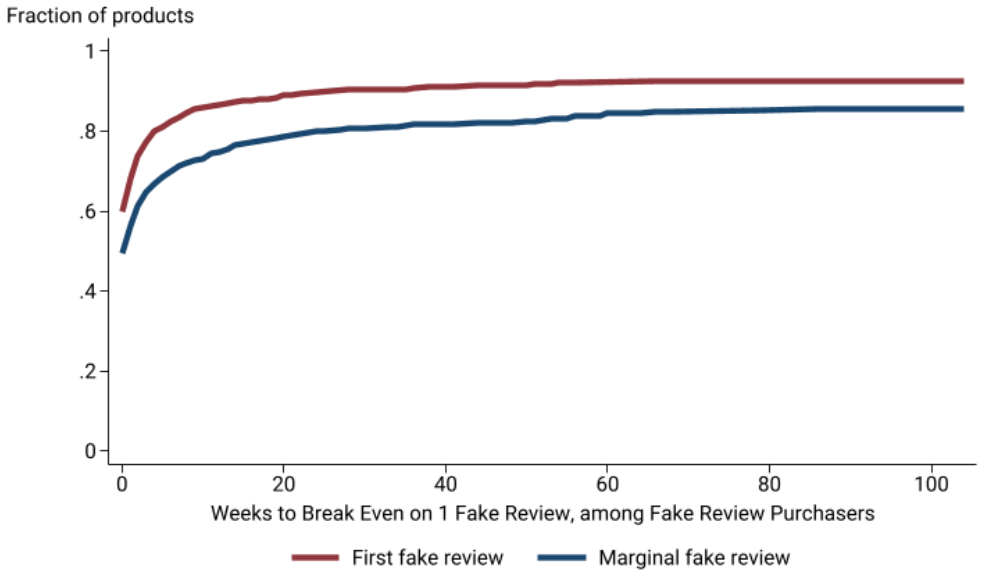
8 Conclusion

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 evi-

Figure 21: Weeks for sellers to break even on one fake review.



(a) First fake review for Fake Review Purchasers and non-purchasers.

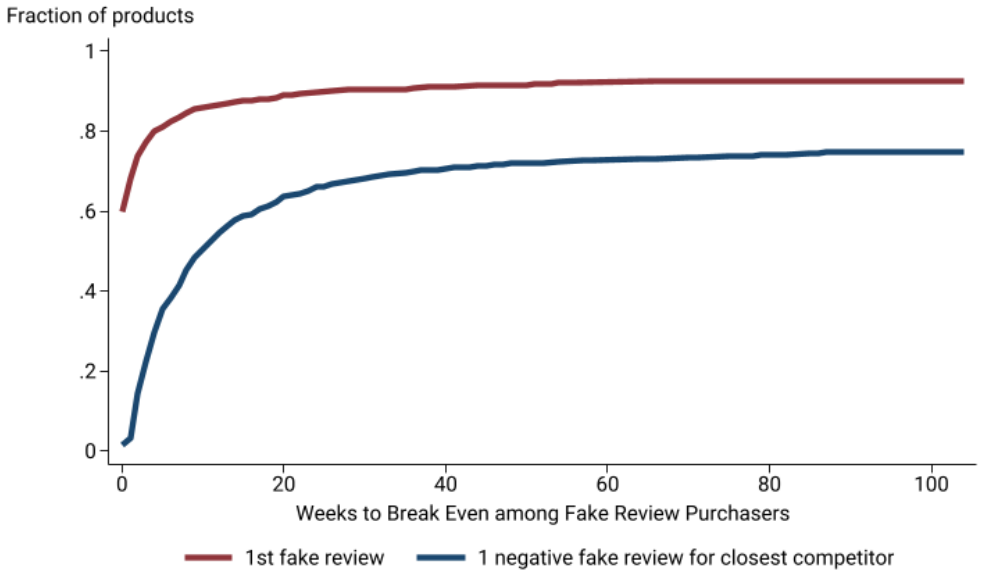


(b) First and additional fake review for Fake Review Purchasers.

dence on the magnitude and nature of consumer harms from this practice in a highly relevant empirical setting, 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

Figure 22: Additional from one negative fake review for closest competitor



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 amelio-

rates 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.

CMA (2020). Cma investigates misleading online reviews. `https://www.gov.uk/government/news/cma-investigates-misleading-online-reviews`. Accessed: 2024-03-18.

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.

FTC (2019). Ftc brings first case challenging fake paid reviews on an independent retail website. `https://www.ftc.gov/news-events/press-releases/2019/02/ftc-brings-first-case-challenging-fake-paid-reviews-independent`. Accessed: 2024-03-18.

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

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.

Johnen, J. and Ng, R. (2024). Harvesting ratings. Technical report, University of Bonn and University of Mannheim, Germany.

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.

Mackay, A. and Miller, N. (2024). Estimating models of supply and demand: Instruments and covariance restrictions. *American Economic Journal: Microeconomics*.

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. (2017). Throwing the Books at Them: Amazon's Puzzling Long Run Pricing Strategy. *Southern Economic Journal*, 83(4):869–885.

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.

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.

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

9 Appendix

9.1 Computing welfare

Since consumers do not observe fake reviews, their expectation of product quality is, in general, different from the econometrician's estimate. Consumers' purchasing decisions are based on the *decision utility* of the product to the consumer, which is computed with consumers' perceived quality, whereas the expected quality that enters the welfare estimation is the *experience utility*, based on the econometrician's estimate of quality having observed fake reviews. 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 23 describes the shape of the distribution of R_j for a given q_j. Figure 24 depicts the joint distribution of (q_j, R_j).

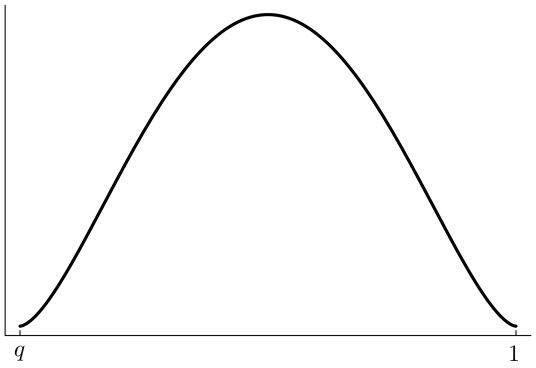


Figure 23: Distribution of R_j with fake reviews.

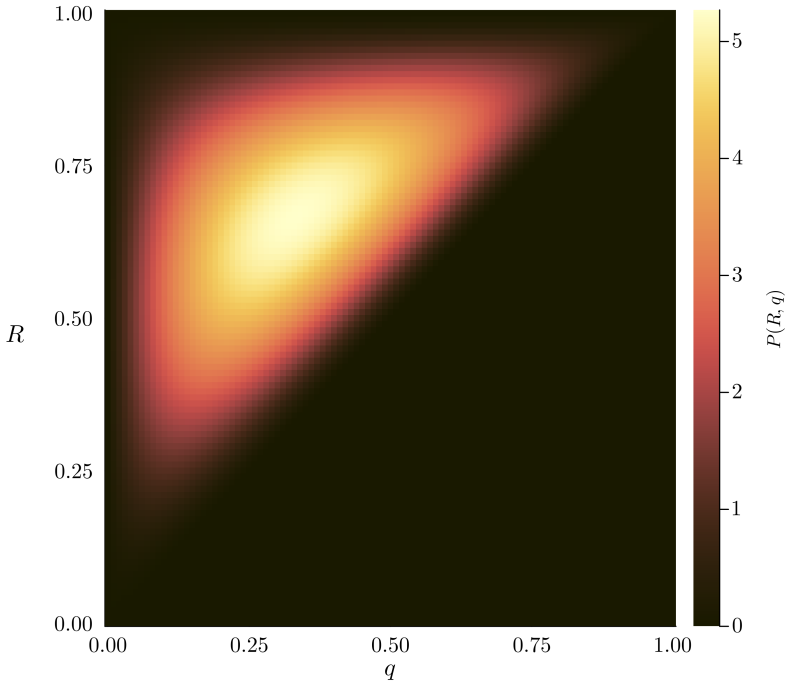


Figure 24: Joint distribution of quality and R

9.3 Posteriors under beta-distributed priors

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

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

For a given product with quality q, the probability that its first N reviews include N^+ good reviews and N^- bad reviews is

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

Given that a product has N reviews split into N^+ positive and N^- negative, the consumer's posterior probability distribution of the product's quality is a beta distribution with parameters $(\alpha_F + N^+, \beta_F + N^-)$:

$$\begin{aligned}
P(q|N, N^+, F) &= \frac{P(N, N^+|q, F)\mu_{Fq}}{P(N^-, N^+|F)} \\
&= \frac{\binom{N}{N^+} p_{Fq}^{N^+}(1 - p_{Fq})^{N^-}\mu_{Fq}}{\sum_{\tilde{q} \in \mathcal{Q}} \binom{N}{N^+} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-}\mu_{F\tilde{q}}} \\
&\approx \frac{p_{Fq}^{N^+}(1 - p_{Fq})^{N^-}\mu_{Fq}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-}\mu_{F\tilde{q}} d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+}(1 - p_{Fq})^{N^-} p_{Fq}^{\alpha_F - 1}(1 - p_{Fq})^{\beta_F - 1} B(\alpha_F, \beta_F)^{-1}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-} p_{F\tilde{q}}^{\alpha_F - 1}(1 - p_{F\tilde{q}})^{\beta_F - 1} B(\alpha_F, \beta_F)^{-1} d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+ + \alpha_F - 1}(1 - p_{F\tilde{q}})^{N^- + \beta_F - 1} d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{B(N^+ + \alpha_F, N^- + \beta_F)}.
\end{aligned}$$

The consumer's unconditional posterior distribution is:

$$P(q|N, N^+) = \sum_{F \in \{0,1\}} \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{B(N^+ + \alpha_F, N^- + \beta_F)} P(F).$$

9.4 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:

$$\log\left(\sum_{q\in\mathcal{Q}} A_{jq}\right) = \log\left(A_{jq'}\right) + \log\left(\sum_{q\in\mathcal{Q}} \exp\left(\log\left(A_{jq}\right) - \log\left(A_{jq'}\right)\right)\right), \tag{16}$$

$$\tag{17}$$

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

$$\log(A_{jq}) = N_j^+\log(p_{F_jq}) + N_j^-\log(1-p_{F_jq}) + \log(\mu_{F_jq}). \tag{18}$$

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

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

9.5 Estimating the Share of Fake Reviews

We discuss in general terms in section 3.1 how we estimate the share of a product's reviews that are fake. In this section we provide more details on this procedure. We rely on the classification model from He et al. (2022a), which 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. For each of the fake review products and their close competitors, for a random sample of roughly 25% of their reviews, we also scraped the pages of the authors of those reviews in order to know the full set of products reviewed by those authors.

We use this data to 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 is measured as a fraction of the subsample of reviewers for which we observe their full rating history. That is, we do not compute the fraction of all reviewers that are designated as fake reviewers, but the fraction of all reviewers with observed ratings histories that are designated as fake reviewers. This provides an estimate of the proportion of fake reviews for that product, but with some noise due to the fact that we only observe ratings histories for a sample of each product's reviewers. For the set products we observe buying fake reviews, the average estimated share of fake reviews is 56% with a median share of 59%.[16]

[16]By contrast, among honest products, we observe only .6% of their reviews are left by these fake reviewers.

10 Survey Appendix

10.1 Review histograms

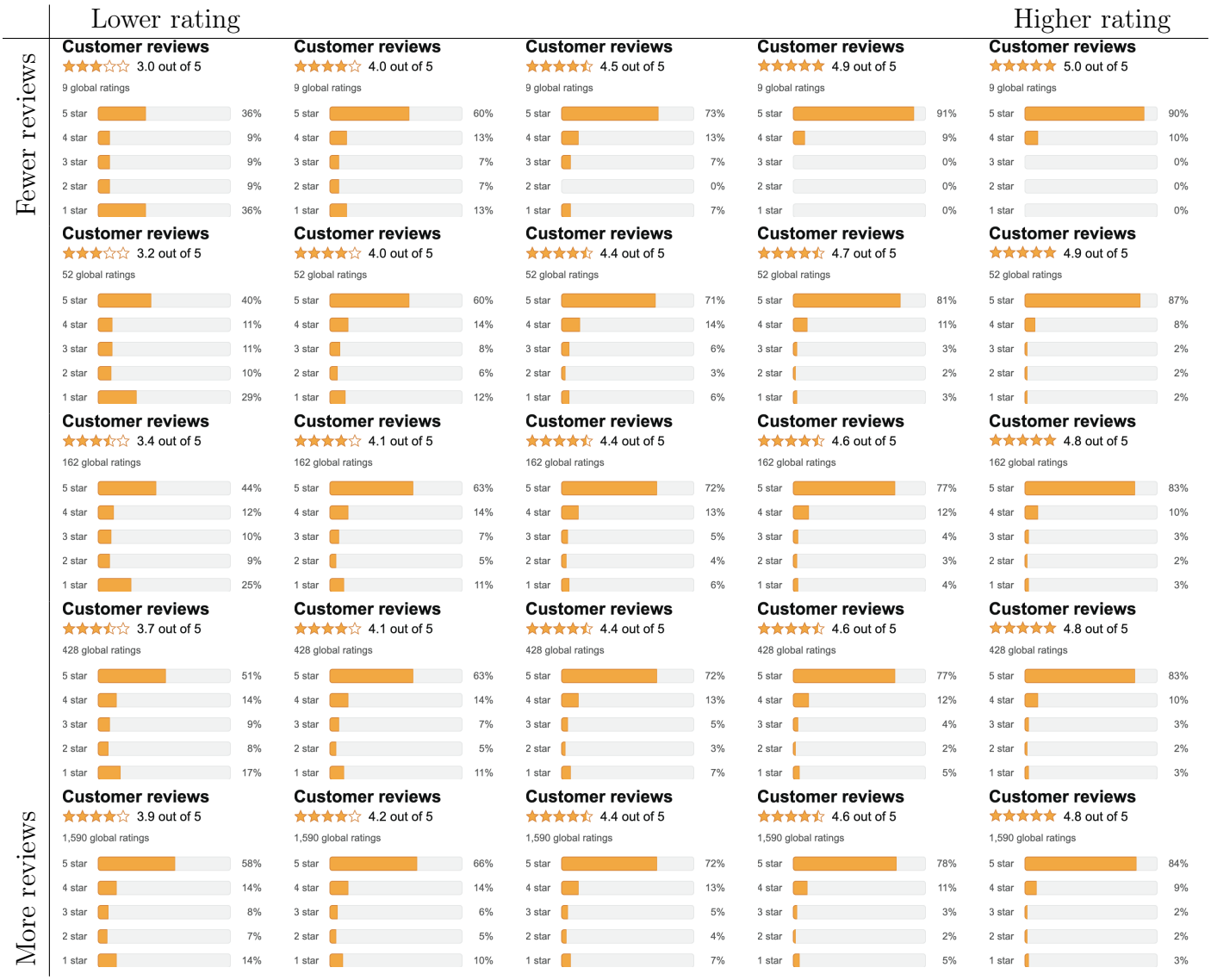


Figure 25: Mean histograms

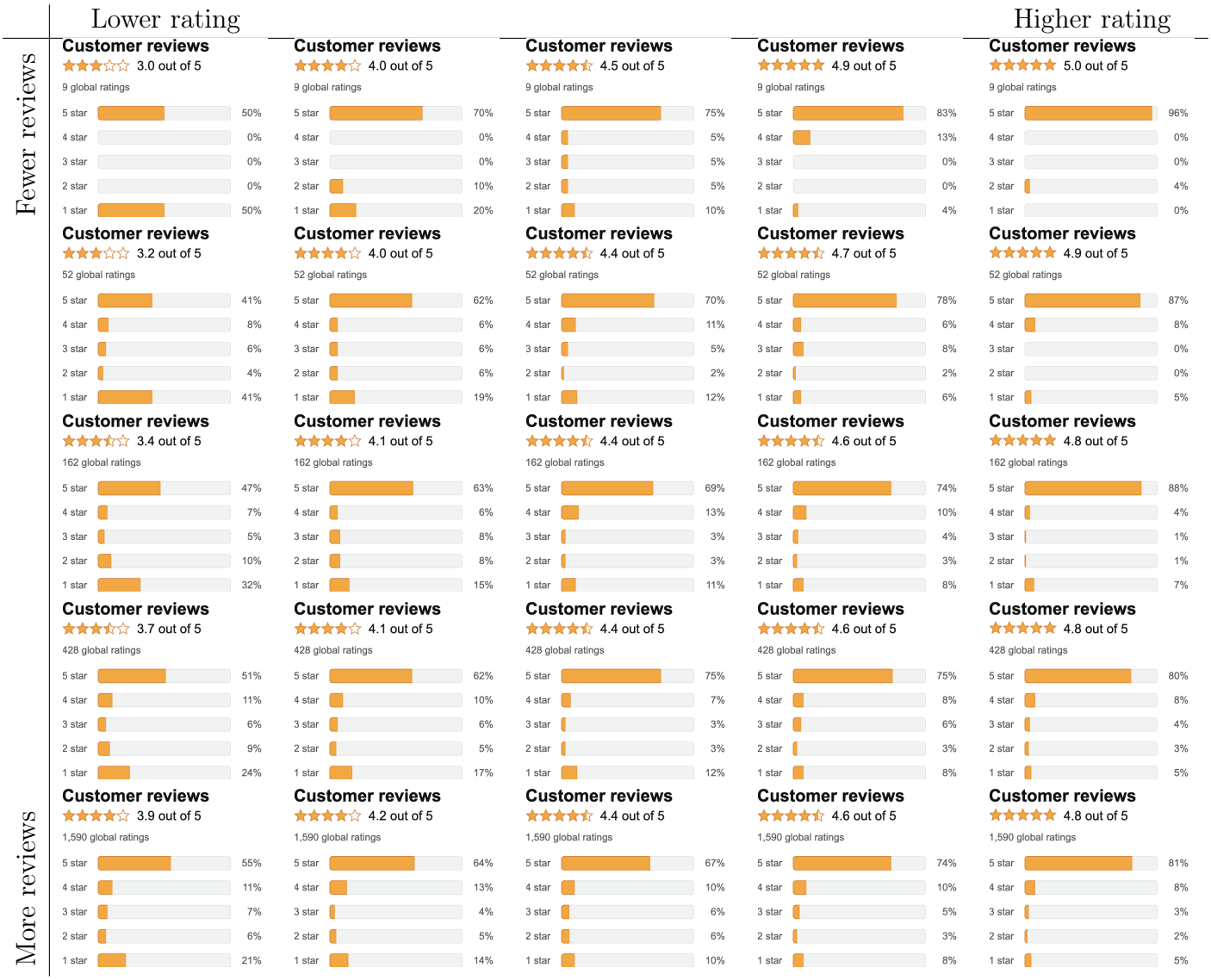


Figure 26: Bimodal histograms

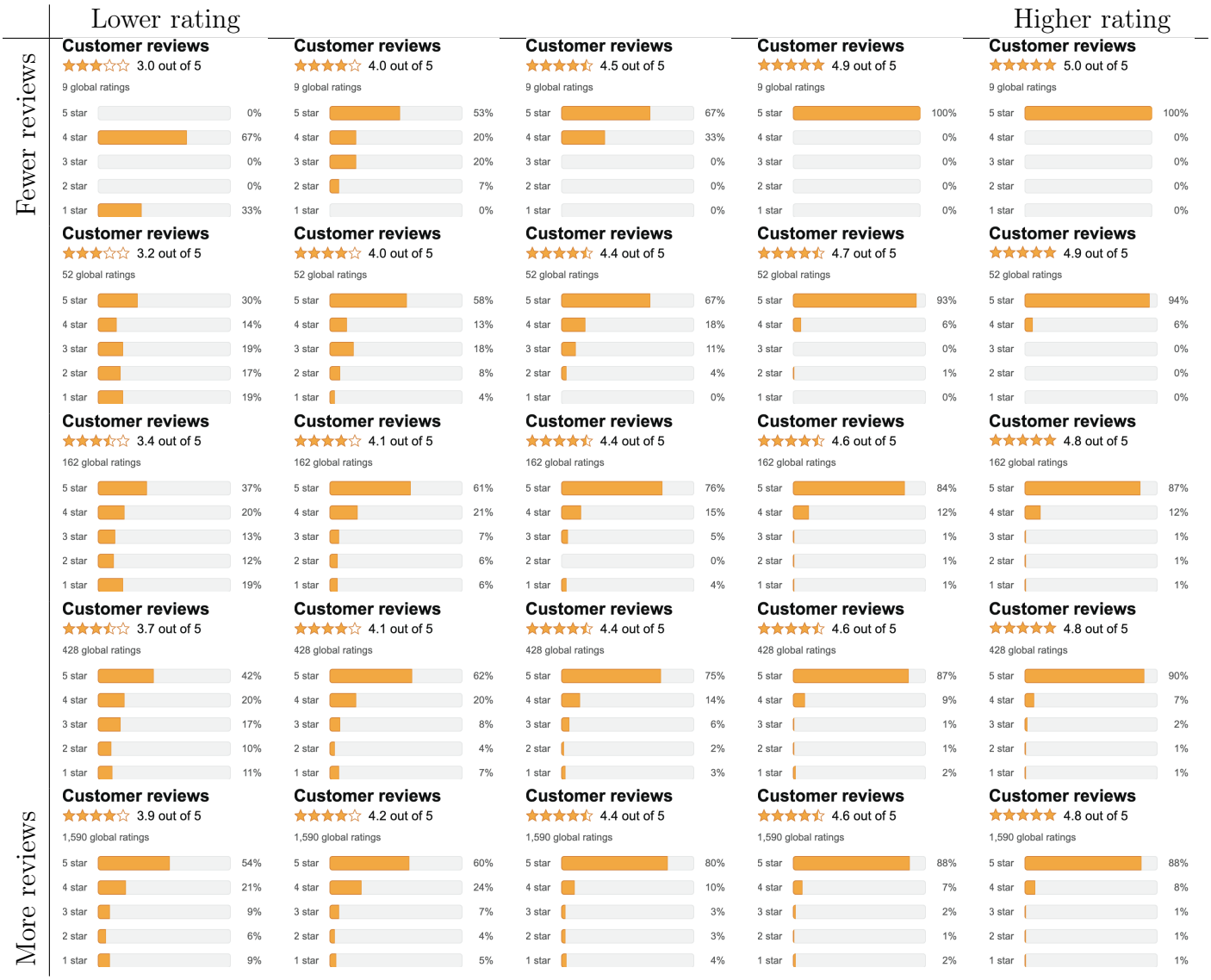
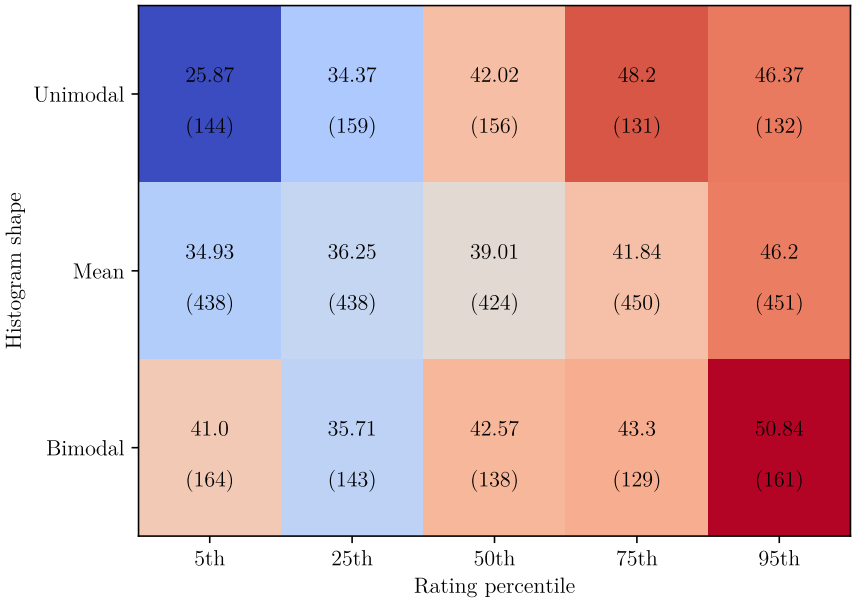
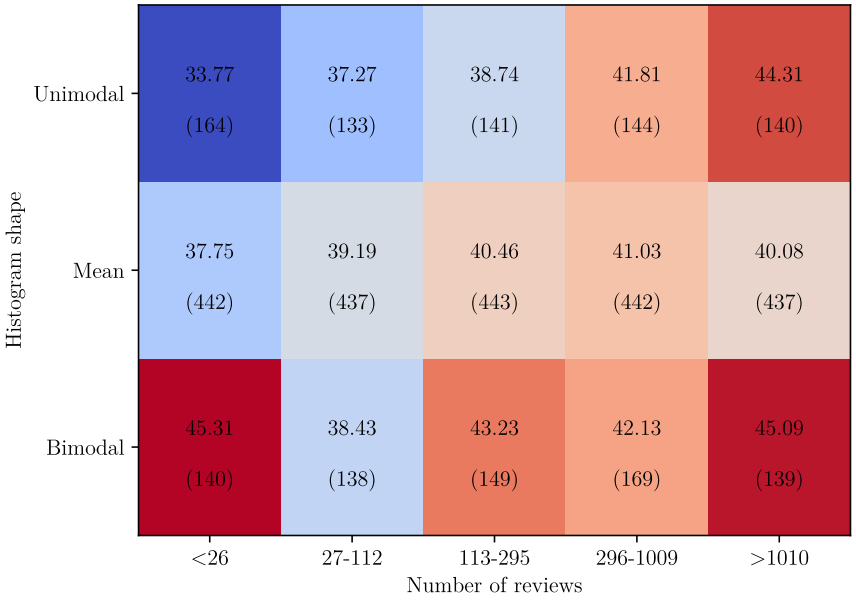


Figure 27: Unimodal histograms


Unimodal
Mean
Bimodal
Histogram shape
25.87
(144)
34.37
(159)
42.02
(156)
48.2
(131)
46.37
(132)
34.93
(438)
36.25
(438)
39.01
(424)
41.84
(450)
46.2
(451)
41.0
(164)
35.71
(143)
42.57
(138)
43.3
(129)
50.84
(161)
5th
25th
50th
75th
95th
Rating percentile


Unimodal
Mean
Bimodal
Histogram shape
33.77
(164)
37.27
(133)
38.74
(141)
41.81
(144)
44.31
(140)
37.75
(442)
39.19
(437)
40.46
(443)
41.03
(442)
40.08
(437)
45.31
(140)
38.43
(138)
43.23
(149)
42.13
(169)
45.09
(139)
<26
27-112
113-295
296-1009
>1010
Number of reviews

10.2 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 5 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 5: Hedonic model of product rank

Log Shares: Lag 1	0.262***	0.281***	0.206***	0.276***
	(18.21)	(18.84)	(15.35)	(18.16)
Log Shares: Lag 2	0.160***	0.177***	0.142***	0.177***
	(11.96)	(12.84)	(10.99)	(12.48)
Log N. Good Reviews: Lag 1	0.100***			
	(14.90)			
Log N. Good Reviews: Lag 2	0.0717***			
	(11.08)			
Cumulative rating: Lag 1		0.0745***		0.0687***
		(9.09)		(8.23)
Cumulative rating: Lag 2		0.0686***		0.0703***
		(8.76)		(8.27)
Weekly rating: Lag 1			0.0232***	0.0200***
			(4.88)	(4.00)
Weekly rating: Lag 2			0.0133**	0.00175
			(2.91)	(0.37)
Log Cumulative N. Reviews: Lag 1		0.105***		0.0900***
		(15.76)		(12.02)
Log Cumulative N. Reviews: Lag 2		0.0758***		0.0595***
		(11.82)		(8.41)
Log Weekly N. Reviews: Lag 1			0.0592***	0.0137
			(8.28)	(1.64)
Log Weekly N. Reviews: Lag 2			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

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

10.3 Deleting Fake Review Purchasers

In this section, we consider the counterfactual policy of deleting Fake Review Purchasers from the platform. We find this policy to be detrimental to consumer welfare in the aggregate. This is true even if we can delete a fraction of Fake Review Purchasers in an ordering that optimizes welfare, as shown in Figure 29. Intuitively, the negative welfare effect arises because any improvement to average quality is outweighed by the price increase from reduced competition in equilibrium. When all Fake Review Purchasers are deleted, the Honest Products are able to set prices that are 1.5% higher and gain profits that are around 15% higher than the factual equilibrium.

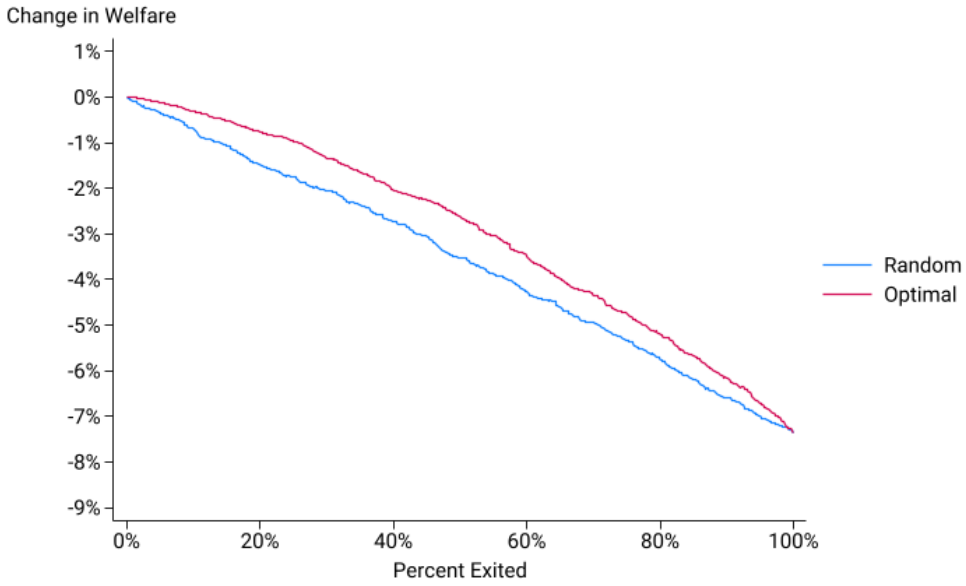


Figure 28: Change in welfare from deletion of Fake Review Purchasers

We also consider the effect of Fake Review Purchasers exiting the market due to lost profits after a counterfactual policy that removes both misinformation and mistrust. We model Fake Review Purchasers exiting according to how much a full deletion policy impacts their profits, and find that exits have an unambiguously negative effect on consumer welfare. The mechanisms governing the equilibrium effects is similar to the deletion counterfactual above, as is the magnitude of the welfare changes.

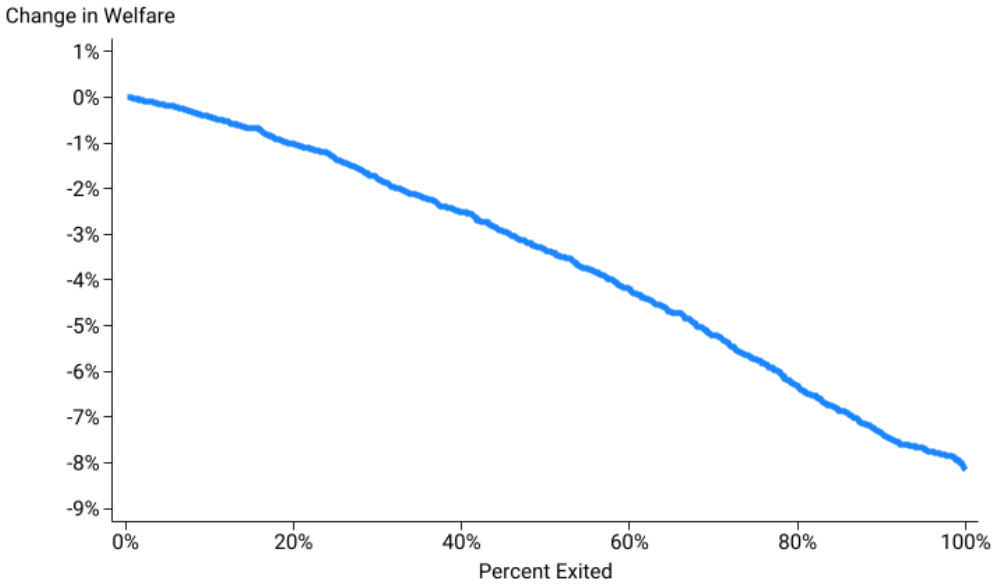


Figure 29: Change in welfare from exit of Fake Review Purchasers