HARVARD | BUSINESS | SCHOOL

The Effect of Price on Firm Reputation

Michael Luca
Oren Reshef

Working Paper 20-128

The Effect of Price on Firm Reputation

Michael Luca
Harvard Business School

Oren Reshef
Washington University in St. Louis

Working Paper 20-128

Copyright © 2020, 2021 by Michael Luca and Oren Reshef.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Pershing Square Fund for Research on the Foundations of Human Behavior and Harvard Business School.

The Effect of Price on Firm Reputation *

Michael Luca †

Harvard Business School

Oren Reshef ‡

Washington University in St. Louis, Olin Business School

Abstract

While a business's reputation can affect its pricing, prices can also affect its reputation. To explore the effect of prices on reputation, we investigate daily data on menu prices and online ratings from a large rating and ordering platform. We find that a price increase of 1% leads to a decrease of 3%-5% in the average rating. Consistent with this, the overall distribution of ratings for cheaper restaurants is similar to that of more expensive restaurants. Finally, these effects don't seem to be driven by consumer retaliation against price changes, but by changes in absolute price levels.

*We are grateful for the comments and suggestions of Dan Elfenbein, Shane Greenstein, Troup Howard, Simon Martin, Jesse Shapiro, Andrei Shleifer, Steve Tadelis, and Reed Walker. Funding for this research was received from the Pershing Square Fund for Research on the Foundations of Human Behavior. We thank Yelp for providing data for this analysis. The right to publish is not tied to the results of this paper.

†mluca@hbs.edu

‡oren@wustl.edu

1 Introduction

A firm's reputation is broadly considered to be one of its most important assets. As a firm's reputation increases, so does the demand for its services and products, and is often accompanied by increases in prices, consumer satisfaction, and profits. At the same time, pricing decisions can also directly affect a business's reputation. For example, when product quality is not immediately observed by customers, increasing prices might signal higher quality. On the other hand, if consumers adjust for price when reporting quality, then a price increase will harm firm's reputation in the long-run. To the best of our knowledge, we are the first to explore the effect of price on reputation, as measured by its online ratings. As elaborated below, our results show a negative relationship between pricing decisions and reputation, highlighting an important downstream consequence of pricing decisions.

We use data from Yelp—a consumer review website for local businesses. We begin by looking at the cross-sectional distribution of Yelp star-ratings for restaurants, broken out by price. If ratings were simply a proxy for quality, one might expect a positive relationship between price and quality—with more expensive restaurants having higher ratings. In contrast, we find that the rating distribution of cheap and most expensive restaurants are broadly similar. The average rating for restaurants in the cheapest Yelp category is 3.4 and the average rating for restaurants in the most expensive category is 3.6—a difference less than a quarter standard deviation. We interpret this as suggestive evidence that ratings are a function of both quality and price.

To shed light on the causal impact of prices, we turn to daily transaction data—looking at orders placed directly through the Yelp platform. We obtain access to item-level information on all food orders finalized on the Yelp Transactions Platform. Our analysis focuses on two main specifications: First, we look for changes in rating in response to changes in prices, controlling for item-week fixed effects. Second, we restrict attention to narrow time bands just before and just after sharp item-level price changes, and estimate the change in online ratings. As an additional robustness test, for a subset of menu items, we take advantage of specific details of our setting to estimate the effect on ratings received for the same item at different price points *at the same time*.

Across specifications, we find that increases in prices lead to lower ratings. A 1% increase in price leads to a 0.05-0.14 decrease in rating on a scale of 1 to 3 (the scale used for delivery purchases), which is approximately 2.5%-5% decrease for the average feedback. This effect becomes increasingly important when considering that the average price change is about 3%-9%. These results are consistent with the cross-sectional evidence, and suggest that higher prices are in fact affecting a restaurant's reputation, and that these effects are both

statistically and economically significant.

We argue that ratings are negatively affected by price-level because consumers adjust their feedback based on prices. We then attempt to disentangle other mechanisms which might affect consumers' rating behavior: We find that the effect is larger and generally more statistically significant for users who are ordering from a restaurant for the first time relative to people who have ordered before. This suggests that consumers indeed respond to price levels rather than use low ratings as punishment for raising prices. This finding also supports the notion that prices are, to some extent, used as a reference point or a signal to set users expectations for quality, as users who have not previously ordered from that business are likely to have the least prior knowledge of quality.

Our results contribute to several streams of literature. First, we contribute to the literature on reputation building. Previous literature has shown that firms can affect their reputation legitimately, by improving quality (Hubbard, 2002, Jin and Leslie, 2009, Board and Meyer-ter Vehn, 2013, Cai et al., 2014, Proserpio and Zervas, 2017) or illegitimately, by faking reviews (Mayzlin et al., 2014, Luca and Zervas, 2016). Our work sheds light on the ways in which pricing decisions influence reputation. Second, our results contribute to the literature on the content of online reviews, and the design of reputation systems (Bolton et al., 2013, Li and Xiao, 2014, Dai et al., 2018, Li et al., 2020). While ratings are often considered a proxy for quality, a growing literature documents biases and issues with consumers' reviewing process (De Langhe et al., 2016, Nosko and Tadelis, 2018, Filippas et al., 2018, Fradkin et al., 2018). Our work shows that ratings account for both quality and price. Our paper contributes to the literature exploring the relationship between prices and ratings (Dodds et al., 1991, Voss et al., 1998, Li and Hitt, 2010). We build on this literature, providing a large-scale empirical analysis of fine-grained observational panel data. More generally, we contribute to the literature on pricing as a tool to signal quality (Klein and Leffler, 1981, Shapiro, 1983, Wolinsky, 1983, Milgrom and Roberts, 1986, Bagwell and Riordan, 1991, Caves and Greene, 1996, Martin and Shelegia, 2019) and benefits of introductory pricing (Cabral et al., 1999, Schlee, 2001).

Finally, our results have implications for businesses, platforms, and customers. Most transaction platforms have some form of rating system to provide consumers with a signal of sellers' quality (for example, Amazon, Taobao, Airbnb). Platforms such as Yelp and TripAdvisor provide online ratings for businesses in offline markets as well. Online ratings have been shown to effect prices and sales (Livingston, 2005, Jin and Kato, 2006, Chevalier and Mayzlin, 2006, Resnick et al., 2006, Cabral and Hortacsu, 2010, Luca, 2016). Our results suggest a tradeoff to increasing prices—in addition to reducing immediate sales, price increases harm firm reputation. This dynamic impacts firms' pricing incentives, and has

material implications to the value of rating to consumers. If consumers are unable to unpack the effect of historic prices on rating, then this puts a wedge between true quality and firm's reputation. Our results also have implications for platform designers, who might be able to mitigate this concern and improve reputation mechanisms by taking into account this effect.

2 Data and Empirical Design

2.1 Settings and Data

We study the effect of price on online ratings in a portion of the food delivery-service industry in the United States, a 35 billion dollar industry, expected to grow at an average annual rate of over 20% in the next 10 years.[1] We focus on Yelp Transactions Platform, an online platform launched in 2013 by the consumers' review website, Yelp. YTP enables users to order from local restaurants.

YTP operates as a part of the standard Yelp website and features a subset of restaurants available on Yelp.[2] Figure 1 depicts the process of ordering on YTP. Figure 1a presents results of a search query on YTP: the shopper views a list of restaurants relevant to the query and location as well as a map of the establishments. Restaurants' data are pulled from the standard Yelp website and include the Yelp Star Rating, number of reviews, food category, and Yelp Dollar Rating. Shoppers can then go to the business page to learn more about the restaurant or initiate an order. Initiating an order redirects users to the restaurant menu page, presented in 1b. Consumers can then choose specific menu items and finalize the transaction.

Following a transaction on YTP, users are prompted with the option to leave a review for the restaurant.[3] Feedback prompt may appear on the Yelp application or via email, depending on the method by which the order was carried out. As displayed in Figure 1c the feedback includes three parts: First, whether the delivery was late, early, or on time. Second, overall experience from the delivery: Great, OK, or Bad (coded: 3, 2, and 1, respectively).[4] Finally, if the consumer had a bad experience, she is solicited to list the specific reasons for the unfavorable experience.[5]

[1] According to a UBS report.

[2] Shoppers are directed to the platform by applying the "Delivery" or "Takeout" filters; by using similar words in a search query; or by initiating an order from the search results page.

[3] In addition, consumers can leave a review on the 'traditional' restaurant yelp page. However, since we cannot directly connect these reviews with specific deliveries, we do not include those in the main analysis.

[4] TThis rating can be compared to the main Yelp Star Rating. We find a correlation of .41 between YTP reviews and Star Ratings when including only restaurants with sufficient reviews (above 20 reviews).

[5] In this paper we think of the review as feedback regrading restaurant's quality. Naturally, the review

We received transaction-level and establishment-level data from Yelp. Data includes all restaurants available on YTP, Yelp's Star Rating, the type of food sold, the business location, and Yelp's Dollar Ratings. Yelp's Star Rating system is a user-generated rating on a one-to five-star scale. Dollar Ratings are meant to approximate the overall cost per dinner, and are assigned by users and aggregated by Yelp. Dollar Ratings are based on users' input, and take on four discrete values: $ = under $10, $$=11-30, $$$=31-60, and $$$$= over $61. We begin by analyzing restaurant-level star-rating from the standard Yelp website. This dataset includes a cross-section of all restaurants in Los Angeles, New York, and Houston (the cities with the largest pool of Yelp reviews), as of January 2019.

For the main analysis, we use proprietary Yelp and YTP data covering a period from YTP's launch in 2013 until January 2019, with the majority of transactions occurring in the last 2-3 years of the data. Here we discuss the main data used for the analysis. A detailed discussion can be found in Appendix B. The data includes all food orders completed on YTP during that period.[6] For each transaction, we observe item-level description and price, the date and time, the identity of the user and business, and, if the user left a review, the ratings given.When looking at orders, we see the price paid—which includes both the base price, and any add-on charges due to item modifications (such as an extra topping on a pizza, or substituting shrimp for chicken in pad thai). Because of this, many observed price differences may not be the result of item-level price changes but consumer-requested item modification. While we cannot always pin down the exact modification, we can generally observe order-item-level description, which sometimes details whether the item included an add-on. The fixed effect regressions, which are used in the second part of the analysis, exclude items with descriptions suggesting that the change in price resulted from a modification rather than a base price change.[7]

Nevertheless, we identify multiple cases where order modifications are not documented in the item description (see Figure A3). To address this issue, we develop a simple algorithm that further restricts our sample to better identify base price changes rather than item modifications. We use the algorithm to focus on reviews received right around price changes. The specific details of the sample selection algorithm are presented in Appendix B. Intuitively,

might also reflect other things, such as the ease of using the online platform or the efficiency of the delivery service. To account for the latter, we include a control for whether a delivery was late. Nevertheless, if consumers are rating anything which is not directly related to the restaurant, then this will likely add noise to our estimate and go against us finding a statistically significant result.

[6] Unfortunately, as part of the agreement with Yelp, we are unable to disclose sensitive business information regarding the *levels* of platform or business performance. We cannot disclose, for instance, the total number of orders or users on the platform, the number of orders per business, or revenue.

[7] For instance, we exclude purchases which include words such as"Adjustment" or "Customize," as well those that explicitly include surcharge amounts.

we begin by restricting attention only to items with sufficient observations and occurrences in which the description does not suggest any item modifications. Then, for each item, we remove price points that appear infrequently in the data (less than five times or 5%) or price points that appear to be changing rapidly, such as \$3, \$5, and then back to \$3. Finally, we exclude periods in which we observe more than one prevailing price. We perform a myriad of robustness tests in order to test the sensitivity of our result to altering the core assumptions of the algorithm, making it more or less conservative (Appendix C).

2.2 Research Design and Empirical Specification

We are interested in identifying the effect of prices on subsequent ratings. The central challenge we face is that prices are endogenously chosen, and thus likely to be correlated with unobserved characteristics such as input costs, quality, or shifts in demand.

One approach is to look at within restaurant variation over time. This might eliminate some of the selection concerns. However, even within restaurants, unobserved demand shocks are likely to impact pricing changes over time. For example, restaurants may become more popular and neighborhoods become trendier. These secular trends however, occur over a long time horizon. Because we have daily data, we can mitigate this concern by focusing on the change in ratings just before and just after a price change – looking for changes in ratings a few days before and after price changes.[8] This essentially rests on the assumption that the exact date that a price change is implemented doesn't coincide with other changes.

We provide several pieces of supporting evidence for this identification strategy. First, we provide suggestive evidence that restaurants don't respond to changes in the competitive environment frequently: over a period of almost 3 years, less than 15% of items ever change prices. The most plausible explanation is that restaurants tend to have relatively high menu costs (Hobijn et al., 2006, Bils and Klenow, 2004). Thus, even if there are sharp changes affecting a restaurant's demand or cost, such as a write up in Food and Wine or a spike in the price of tomatoes, menu costs create a friction that hinders updating prices in the very short run. Second, we find that prices usually change at the restaurant level rather than for specific items. For instance, Figure A1 shows the number of price changes in a given week within a restaurant for the three restaurants with the most price changes in our data. It is clear that restaurants rarely update specific items, but instead redesign the whole menu. In fact, the median restaurant changes about one third of its menu each time it updates prices. This suggests that restaurants wait until they have a sufficient number of changes to

[8] In the main specification, we use 10 days before and after the price change. In Appendix C, we test the robustness of the main results to alternative definitions.

justify the costly price update. Thus, price changes are not correlated, at least in the very short run, with item-level changes, such as changes in dish recipe or ingredients, but with restaurant-level changes.

Some changes, such as kitchen renovation or hiring a new chef, might affect all dishes simultaneously. However, it seems unlikely that price would adjust instantly (over a few day period) as one can expect a transition period for these improvements to affect food quality. Moreover, one might expect genuine improvements in quality to tend to lead to higher prices and higher ratings. This goes in the opposite direction of the reputational harm of higher prices that we observe.

Finally, as a robustness test, we estimate an additional specification, which takes advantage of an institutional detail unique to YTP. Since YTP operates with multiple partners, who sometimes tend to update prices at varying quickness, we can occasionally observe the same item being sold at different prices *at the same time*. This allows us to isolate the effect of price on ratings, while holding fixed all other factors that might affect restaurant's ratings.

The first empirical specification uses a myriad of fixed effects to estimate the effect of price on ratings, controlling for item, restaurant, and time effects:

$$Y_{jt} = ln(Price_{jt}) + X_{jt} + \gamma_w + \delta_j + \mu_{wj} + \epsilon_{jt} \tag{1}$$

An observation is one item in a transaction. w is the week index, j denotes item-business combination , and t is an index for specific dates. Y denotes the outcome of interest: Order rating, on a scale of 1 to 3, or whether the order received (did not received) the highest (lowest) ratings. $Price$ is the price of the item per order, excluding taxes and delivery fees.[9] X is a vector of controls which includes: a dummy for pickup or delivery, the share (in monetary terms) of the item out of the total transaction price, whether the delivery was marked as 'late', the delivery partner, and day-of-the-week (non-parametrically). γ_w denotes calendar week fixed effect, δ_j denotes item-level fixed effect, and μ_{wj} denotes week-item fixed effect. Naturally, when the latter is included, it absorbs the first two fixed effects. To be conservative, standard errors are clustered at the item-business level. The parameter of interest, β, should be interpreted as the impact of percentage changes in price on the outcome variable. This specification uses the full sample, excluding items with descriptions suggesting item modifications.

[9] In our data, ratings are given at the order-level, while prices are determined at the item-level. This is likely to add noise to our estimate and works against rejecting the null hypothesis. Nevertheless, in Appendix C we conduct robustness checks in which we restrict our analysis only to the largest (most expensive) item in each order.

The second empirical specification focuses on narrow time windows around sharp price changes:

$$Y_{jt} = ln(Price_{jt}) + X_{jt} + \lambda_G + \epsilon_{jt} \tag{2}$$

Here—beyond the indices and variables defined in equation 1—G is an index for item-business X price change and λ is the fixed effect per item price change.[10] Standard errors are clustered at the item-price level. This specification uses the most restrictive sample selection criteria, based on the algorithm discussed in Section 2.1 and Appendix B.

3 Results

3.1 Motivating Evidence

To motivate the main analysis we start by analyzing the relationship between prices and ratings in the standard Yelp website. We do not have the exact prices for this broader set of restaurants. However, we have a price index that Yelp uses, which is in terms of Yelp Dollar Signs (as described in Section 2.1). We use a cross-section from January 2019 on all restaurants in New York City, Los Angeles, and Houston, the three cities with most restaurants on Yelp.

Figure 2 presents the distribution of ratings (at the restaurant level) broken out by price. Panel A presents the raw distribution of Star Ratings by Dollar Ratings, and Panel B presents the same distribution controlling for city and food type, i.e. pizza, Mexican food, etc. The figure suggests that, though higher prices are correlated with higher ratings, the effect is marginal. The distributions mostly overlap, with similar modes but slightly fatter tails at the lower end for cheaper places.

Appendix Table A1 presents the results formally. In general, we find no statistically significant differences between the ratings of one- and two-dollar restaurants. We find that three-dollar restaurant received higher ratings than one-dollar restaurant, but the relation is economically small: Across all specifications, on average, three-dollar restaurants have 0.14 higher star ratings compared to one-dollar restaurants (an increase of about 3.5% for the median business),[11] even though, on average, they are more than 4 times more expensive.

[10] Consumers may respond differently to price increases and price reductions. Unfortunately, we are unable to separate the two since we rarely observe any price reductions. Thus, our results should be generally interpreted as the ratings' response to price increases

[11] Note that in this specification we look at the standard Yelp Star Rating, which is given on a scale of 1 through 5. In contrast the main analysis is based on the YTP order ratings, which are on a 1 to 3 scale.

This suggests that more expensive restaurants do not receive overwhelmingly higher ratings, consistent with our hypothesis that ratings are price adjusted. One alternative explanation is that more expensive restaurants offer the same level of food quality as cheaper ones. If that is indeed the case, however, it would be difficult to explain how these expensive restaurants manage to stay in business over time without offering compensation for the hefty price tag. Nevertheless, in the next subsection we attempt to more directly identify the relation between price and ratings.

3.2 Evidence from YTP

Subsequent analysis focuses on ratings received on YTP. The main advantage is that this data allows us to match reviews with orders, specific items, and prices. In particular, we use price variation within items to estimate the effects of price changes on ratings received.

Fixed Effects Regressions The first set of results uses a myriad of fixed effects to control for unobserved differences across items and within-item over time. The results are presented in Table 2. Each column presents an estimation of Equation 1, with a different set of fixed effects. The specific fixed effects are detailed below each column.

Column (1) presents the effect of price on ratings without any item- or week-level fixed effects. While statistically significant, the estimated relation is economically small, a 1% increase in price leads to an average increase of 0.006 in rating, on a scale of 1 to 3. The average price change in this sample is about 3%, which implies an increase of about 0.018 in rating, or about 0.006% change. This positive but weak result echoes the pattern presented in Figure 2, and suggests that higher priced items are not much higher rated in the cross-section.

Column (2) adds item- and week-level fixed effects. The estimated effect of prices on subsequent ratings is statistically significant and *negative*. We find that a 1% increase in prices causes an average drop of 0.052 in ratings. The same back-of-the-envelope calculation as conducted above, suggests that for the average price changes, this implies a decrease of about 5.6% in restaurant's ratings. Column (3) presents a similar result, with an even more restricting specification, which accounts for the item-week fixed effect, effectively comparing ratings of the same item at different price points within a given calendar week. The coefficient is -0.048 and is statistically significant.

Columns (4) and (5) decompose the three levels, by examining the effect of price on the (linear) probability of receiving the highest rating ("Great") or *not* receiving the lowest

Thus, similar magnitudes in absolute terms have different interpretations in percentage terms.

rating ("Bad"), respectively. We find that prices affect ratings across the board; higher prices both decrease the likelihood of receiving the highest ratings (Column (4)) and decrease the likelihood of *not* getting the lowest rating (Column (5)).

In Appendix C, we conduct several robustness test to these results: As an additional robustness check, we conduct an analysis in which we restrict attention only to the largest or most expensive item in each order Additionally, we test the sensitivity of the analysis to different levels of clustering the standard errors, such as at the order, business, or user-level. Finally, we test whether the effect persists when controlling for reviewer identity, i.e., we compare reviews for the same item sold at a different price to the *same* consumer.

Taken together, the results of Table 2 find that, within-item and in a small time frame, prices have a negative effect on rating. This result is consistent with our main hypothesis; conditional on quality, as firms raise prices, subsequent ratings suffer.

Sharp Price Changes For the following analysis, we restrict attention to reviews given around sharp price changes. Our sample selection criteria is intendedly conservative, and thus we omit the vast majority of data and remain only with a small core of cleanly defined price changes. A detailed discussion on the rationale guiding our decisions is presented in Appendix B.

Figure 3 presents the results graphically; We find a sharp drop in average ratings following a price increase. The formal results are presented in Table 2. Due to the restrictive sample selection rules we impose, the number of observations and items is substantially lower than present in Table 1. This approach, while reducing our sample size, provides arguably the cleanest estimate of the effect of prices on restaurant rating. Examining only sharp price changes allows us to avoid issues related to spurious price changes and incomplete data.

Column (1) presents the main specification. We find that a 1% increase in price leads to an average decrease of 0.11 in rating on a scale of 1 to 3. For this sample, the average price change is about 9%, suggesting that an average price change is decreasing subsequent ratings by over 30%. This effect is substantially larger than described in Table 2 because the estimated coefficient is more than double in magnitude, and because the average price change in this sample is almost triple in size. The latter is likely to be the result of the sample selection criteria, which might only be picking up large price changes, and excluding small price changes from the analysis. Nevertheless, even for the average price change in the sample, 3%, the predicted reduction in ratings is about 12%.

In Columns (2) and (3) we change the selection criteria, making it either more lenient or more restrictive, respectively. A detailed discussion of how these alternative samples are

constructed is presented in Appendix B.[12] The estimated magnitudes are consistent with the ones presented in Column (1) and remain statistically significant across specifications. Columns (4) and (5) estimate the effect of price increases on the probability of receiving the highest rating or not receiving the lowest rating, respectively. The estimation uses a linear probability model. Together, these two columns suggest that the effect is primarily driven by reduction in the highest score, with about 66% attributed to reduction in the highest ratings and about 33% from increases in the lowest rating category.

We conduct several robustness tests to explore the sensitivity of the main results to variable definitions and coding. We also examine the importance of the allowed window around price changes.[13] In addition, we relax the linear probability assumption using ordered and conditional logit models. Finally, similar to the fixed effects regressions, we test the robustness of our finding to including only the most expensive item in each order, and the sensitivity to different levels of clustering the standard errors. Appendix C presents the estimation results of robustness tests.

Across the board, we find evidence that price increases lead to reductions in ratings. The estimated effect is consistent across specifications and is around 0.11, i.e. a 10% increase in price leads to a 1.1 average increase in rating (over 30%). We thus conclude that prices have a significant and substantial negative effect on user ratings.

3.3 Mechanism and Alternative Research Designs

The Role of Retaliation In the previous section we found that price has a deleterious effect on subsequent firms' ratings. In this section, we explore whether the effect is driven by the price levels, or by existing customers who are retaliating against the change in price. For instance, Kahneman et al. (1991) and Eyster et al. (2017) highlight the ways in which reference points can affect decisions. Repeat consumers might think about the restaurant's previous prices as a reference point, and retaliate against price increases. Alternatively, consumers might simply be reacting to the price expressing their discontent with the price change (Hirschman, 1970).

To shed light on this, we split the sample into existing customers (who have seen both

[12] To summarize: First, there are cases in which there is some overlap between two prevailing prices. In the main specification we exclude all of these observations. In contrast, in Column (2) we use the lowest price. This definition is somewhat more lenient than the main specification. Second, Column (3) presents the most restrictive specification, we only include products that ever had two prices and that these prices never overlap.

[13] Consistent with Column 1 in Tables 1 and A1, the effect diminishes as the window becomes larger, but becomes larger in magnitude when we narrow the window.

the lower and the higher price) and new customers (for whom prior menu prices are less salient). If the effect is driven by retaliation against price changes, one would expect the effect to attenuate for users who haven't ordered from the specific restaurant in the past.

Table 3 shows the effect of prices on ratings when restricting the sample only to users who order from a specific restaurant for the first time. The structure is similar to Table 2, but with the sample restricted to new users. Column (1) presents the estimation results for our main specification. The coefficient is -0.126, is slightly larger in magnitude than the one estimated in the full sample, -0.11. Similarly, the estimates in Columns (2) - (5) are slightly larger than their counterparts using the full sample.[14]

These results suggest that retaliation against price increases is not the main driver of the negative relation between prices and ratings, as first-time users seem to be more responsive to price increases. We interpret these as suggestive evidence of the role ex ante expectations in determining ratings. The fact that new consumers are more affected by price together with the fact that price may signal higher quality, can be indicative of the fact that consumers are reducing their rating because the item does not live up to the expectations formed based on its hefty price tag.

Alternative designs The main identifying assumption of the main research design is that within a narrow time window price changes are orthogonal to other unobserved changes in items. While we provide some anecdotal evidence to support this claim, this assumption is fundamentally untestable. In order to corroborate our main results, we introduce two alternative research designs.

First, as discussed in Section 2.1 and presented in Figure A3, for most items there are no price changes and when there are price changes, they occur simultaneously. We interpret this stylized fact as implying that price changes are either stemming from restaurant-level shocks or that restaurants have high menu costs and update their menu only when a sufficient number of changes has accumulated. Either way, we argue that restricting our attention only to times when multiple prices were simultaneously adjusted can, at least partially, alleviate some of the potential concerns regarding item-level unobserved changes.

To this end, we conduct additional analysis, restricting attention to weeks in which we observe more than 5, 10, or 20 of the restaurant's items changing price; a sizable change given that the average restaurant in our sample only sells about 30 items. The results are presented in Appendix Table A7. Similar to the main analysis, we find a consistent negative and significant relation between price and ratings. The estimated magnitudes are

[14] In Column (3) the estimate is much larger and noisier than the other specifications. This might be due to the sharp decrease in the number of items included in that sample. We can not, however, reject the null that it is the same as the coefficient estimated in the main specification, -0.135.

substantially larger than the main specification. However, the results are much noisier, probably due to the smaller number of observations.

Second, while the above design helps alleviate price changes related to item-level unobserved changes, one potential concern that remains is that all restaurant items change simultaneously. Ideally, we would want to compare the exact same item at the same time, but at different prices. As it turns out, for a small subset of items, we can actually see that natural experiment. To understand how this might occur, we need to understand the institutional details behind YTP: YTP does not perform the delivery itself but connects users with delivery firms, referred to as partners. If the same restaurant is affiliated with multiple partners, then the YTP algorithm assigns the delivery to one of the partners. Notably, consumers cannot change that decision. Apparently, there are multiple instances in which an item changes price, but certain partners are quicker than others to update the price.[15] Thus, there are short time windows in which different partners sell the same item at different prices, one at the old price and one at the new price. An example of an actual item can be found in Figure A2; The price increased from $7 to $9 for both partners, but partner B updated the price before partner A did. This means that, for a short time window, the same item sold at both $7 and $9. Thus, we can treat prices as randomly assigned and estimate the effect of price increases on ratings.

The results are presented in Table A8. Column (1) presents the estimation results on rating. The coefficient on log price is -0.075. This estimate is slightly smaller than the main specification, though we cannot reject the null that the estimated coefficient is equal to the average result from the main specification, -0.11. Columns (2) and (3) present the decomposition by highest and lowest rating received. Similar to the main specification, the magnitude of the effect on the number of "Great" reviews is larger. This effect, however, as opposed to the effect on the number of reviews that are not "Bad" is not statistically significant, possibly due to the small number of observations.

Reviewers Selection Price changes also have the potential to affect the type of consumers going to and reviewing a restaurant. For example, customers who tend to visit restaurants with higher prices might be more (or less) stringent reviewers than those who go to restaurants with lower prices. To explore whether selection of reviewers is driving the relation between price and ratings, we construct a harshness score for each user based on the average rating given across all transactions excluding the current transaction (leave-out mean), as described in Appendix B. We first test whether price levels are correlated with

[15] It doesn't seem that there are particular partners which are always late and others that are predominantly early to update price. In any case, we control for partner identity in all specifications.

reviewers harshness, and second whether controlling for reviewers' harshness mitigated the relation. The results are presented in Table A6. We do not find evidence of significant selection of harsher reviewers following price increases. For instance, in Column (1) we see that the effect of price on average reviewers' harshness is small and statistically insignificant. Similarly, we find that controlling for reviewers' harshness slightly decreases the estimated coefficients, but the results remain statistically and economically significant.

Another possibility is that reviewers who are less price sensitive or have strong preference for a specific item go for items with higher prices. Our main results show that reviewers adjust their reviews for price. To the extent that reviewers of higher priced items are less price sensitive or have strong preferences, the adjustment we see is likely smaller than if reviewers were randomly drawn.

4 Discussion

In this paper we study the effect of price on firm reputation, as measured by its online ratings. We collaborate with Yelp Transaction Platform to obtain item-level prices and ratings. Using several research designs, we find that price increases lead to a decrease in subsequent ratings. Our preferred specification suggests that a 1% increase in price leads to a 0.11 decrease in ratings, about 4% for the average restaurant. Thus, *ratings are price-adjusted* rather than conveying objective quality.

This result has several important implications to consumers, firms, and rating system designers. First, the results speak to the welfare gain and value of reputation systems to consumers. For instance, if consumers are unable to unpack the effect of historic prices on ratings, or they have some incorrect beliefs about how raters incorporated price into reviews, then this would put a wedge between items' true quality and the perceived reputation. This mechanism does more than introducing noise into consumers' decision-making process, but instead creates consistent biases in specific directions.

Second, consistent with the above analysis, strategic sellers might be tempted to take advantage of misguided consumers by setting low introductory prices. Initial low prices can mechanically boost ratings and allow some firms to eventually take advantage of their good reputation by increasing sales and prices. More generally, our results point to a tradeoff—price increases don't just reduce present demand, but can potentially harm future demand by decreasing firm reputation. We expect this negative effect to attenuate as the number of reviews grows larger.[16] Finally, platform makers can improve existing reputation mechanisms

[16] Our work focuses on ratings for delivery orders. The effect is likely to vary across contexts - across

by redesigning the rating mechanism to account for the effect of historic prices on reviews received.

Lastly, in this setting the mechanism driving the adverse impact of prices on ratings is somewhat unclear. We believe the two most plausible explanations are that consumers rate net utility, i.e. quality minus (or over) price, rather than quality, or that consumer rate the deviation from their ex ante expectation, i.e. value minus expectation, and that expectation is positively correlated with prices. Unfortunately, we are unable to fully disentangle these to mechanisms in our setting. We believe that the effect is driven by a mixture of these two forces: On the one hand, we do find that repeating users are affected by price, which implies that deviations from expectations are not the sole mechanism driving the results. On the other hand, new consumers seem to be more affected, so expectations seem to play some role. Future work might be able to take advantage of other settings and institutional details in order to decompose the main effect and quantify the relative magnitudes of these two forces.

platforms (e.g. Yelp vs UberEats), across categories (e.g. products vs services), and across user types (e.g. in restaurant diners may react differently than delivery orders). One direction for future research is to explore the extent to which this effect generalizes to products, other services, and even dine-in restaurants.

References

Bagwell, K. and Riordan, M. H. (1991). High and declining prices signal product quality. *The American Economic Review*, pages 224–239.

Bils, M. and Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5):947–985.

Board, S. and Meyer-ter Vehn, M. (2013). Reputation for quality. *Econometrica*, 81(6):2381–2462.

Bolton, G., Greiner, B., and Ockenfels, A. (2013). Engineering trust: reciprocity in the production of reputation information. *Management science*, 59(2):265–285.

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

Cabral, L. M., Salant, D. J., and Woroch, G. A. (1999). Monopoly pricing with network externalities. *International Journal of Industrial Organization*, 17(2):199–214.

Cai, H., Jin, G. Z., Liu, C., and Zhou, L.-a. (2014). Seller reputation: From word-of-mouth to centralized feedback. *International Journal of Industrial Organization*, 34:51–65.

Caves, R. E. and Greene, D. P. (1996). Brands' quality levels, prices, and advertising outlays: empirical evidence on signals and information costs. *International Journal of Industrial Organization*, 14(1):29–52.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

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

De Langhe, B., Fernbach, P. M., and Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6):817–833.

Dodds, W. B., Monroe, K. B., and Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. *Journal of marketing research*, 28(3):307–319.

Eyster, E., Madarasz, K., and Michaillat, P. (2017). Pricing when customers care about fairness but misinfer markups. Technical report, National Bureau of Economic Research.

Filippas, A., Horton, J., and Golden, J. (2018). Reputation inflation. *New York University, Working Paper*.

Fradkin, A., Grewal, E., and Holtz, D. (2018). The determinants of online review informativeness: Evidence from field experiments on airbnb.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, volume 25. Harvard university press.

Hobijn, B., Ravenna, F., and Tambalotti, A. (2006). Menu costs at work: restaurant prices and the introduction of the euro. *The Quarterly Journal of Economics*, 121(3):1103–1131.

Hubbard, T. N. (2002). How do consumers motivate experts? reputational incentives in an auto repair market. *The Journal of Law and Economics*, 45(2):437–468.

Jin, G. Z. and Kato, A. (2006). Price, quality, and reputation: Evidence from an online field experiment. *The RAND Journal of Economics*, 37(4):983–1005.

Jin, G. Z. and Leslie, P. (2009). Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 1(1):237–67.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of political Economy*, 89(4):615–641.

Li, L. and Xiao, E. (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science*, 60(8):2054–2072.

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

Li, X. and Hitt, L. M. (2010). Price effects in online product reviews: An analytical model and empirical analysis. *MIS quarterly*, pages 809–831.

Livingston, J. A. (2005). How valuable is a good reputation? a sample selection model of internet auctions. *Review of Economics and Statistics*, 87(3):453–465.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

Martin, S. and Shelegia, S. (2019). Underpromise and overdeliver?–online product reviews and firm pricing. *Working Paper*.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of political economy*, 94(4):796–821.

Nosko, C. and Tadelis, S. (2018). The limits of reputation in platform markets: An empirical analysis and field experiment. Technical report, National Bureau of Economic Research.

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5):645–665.

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental economics*, 9(2):79–101.

Schlee, E. E. (2001). Buyer experimentation and introductory pricing. *Journal of Economic Behavior & Organization*, 44(3):347–362.

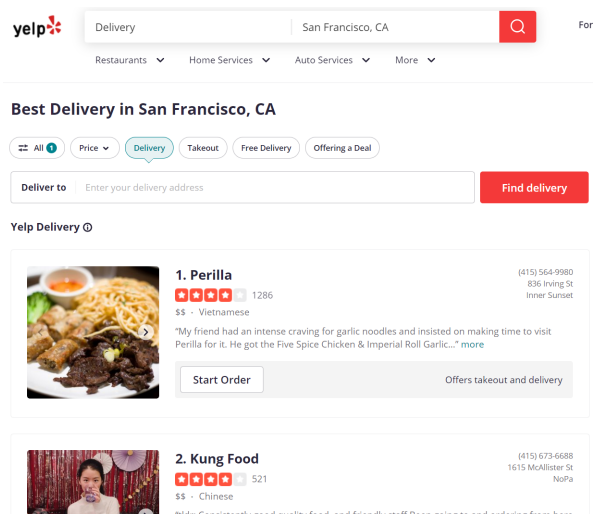
Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The quarterly journal of economics*, 98(4):659–679.

Voss, G. B., Parasuraman, A., and Grewal, D. (1998). The roles of price, performance, and expectations in determining satisfaction in service exchanges. *Journal of marketing*, 62(4):46–61.

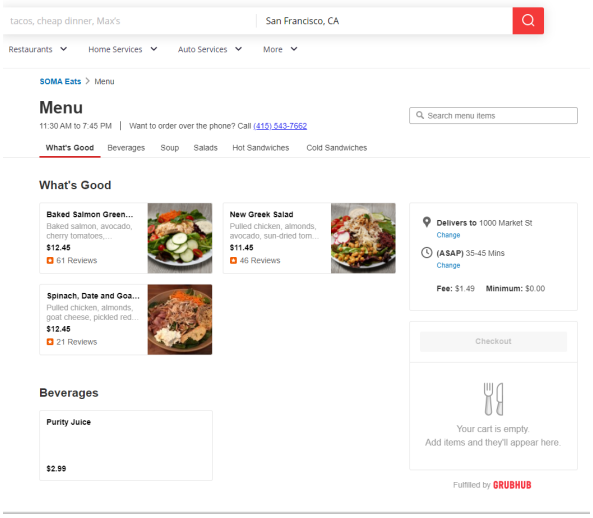
Wolinsky, A. (1983). Prices as signals of product quality. *The review of economic studies*, 50(4):647–658.

Figures

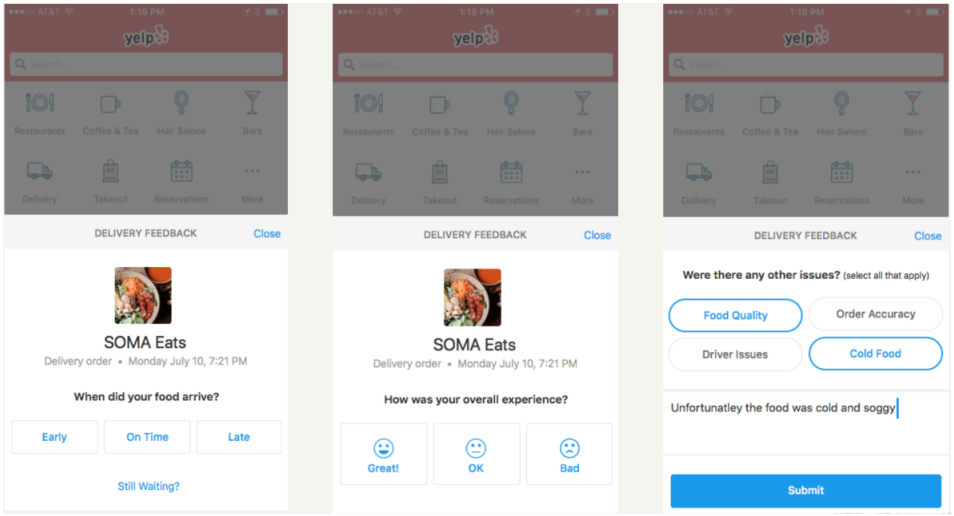
Figure 1: Visualization of the Ordering and Review Process on Yelp Transactions Platform



(a) Searching on YTP



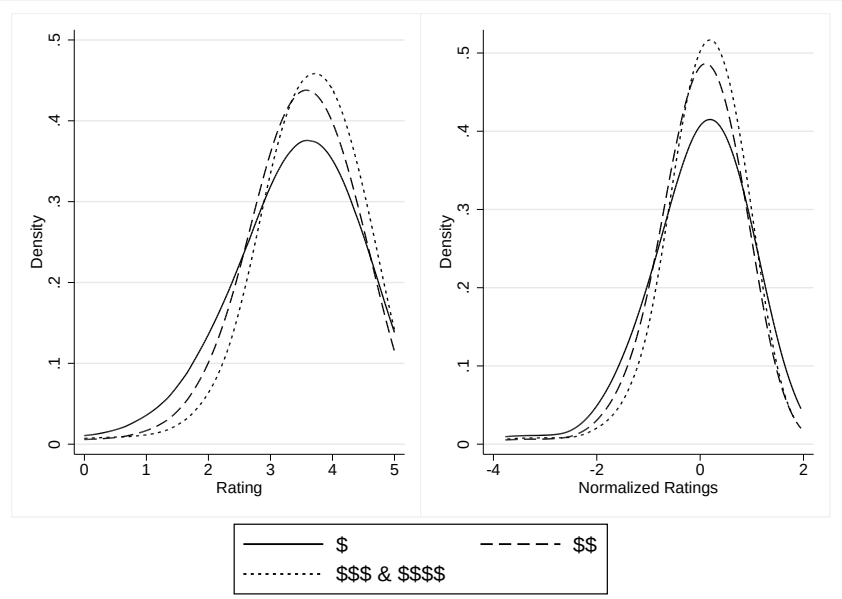
(b) Ordering on YTP



(c) Rating on YTP

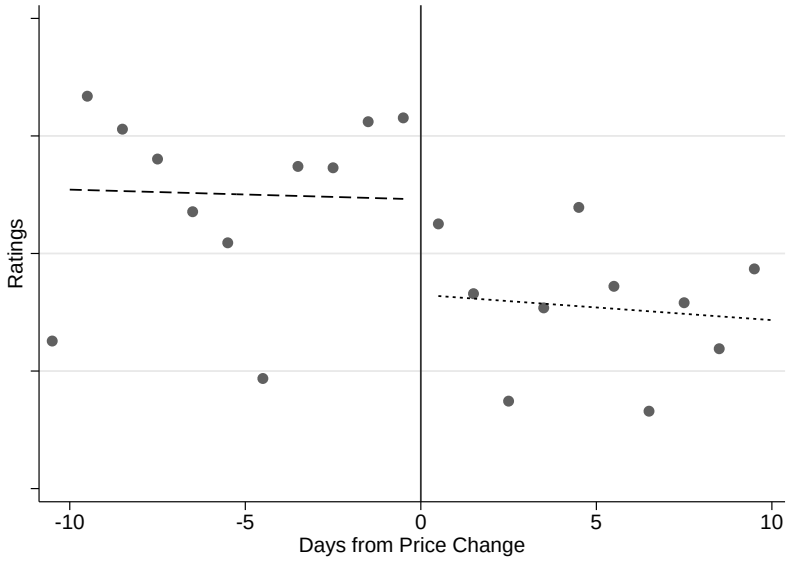
Note: Panel A presents the search results for delivery on YTP in San Francisco. Panel B presents the menu where consumers can pick the specific items and finalize the transaction. Panel C presents review process on YTP.

Figure 2: Motivating Evidence from a Cross-Section of Yelp Star-Ratings



Note: This figure presents the density distribution of Yelp Star-Ratings by Dollar-Ratings in the three cities with the most restaurants on Yelp, as of January 2019. Panel A present the raw distribution. Panel B presents the normalize distribution by city and food category.

Figure 3: The Effect of Price Change on Ratings



Note: This figure presents the raw relation between ratings and days to price increases for the items with the most meaningful price changes. The day of price increase is normalized to zero.

Tables

Table 1: The Effect of Prices on Ratings Using Fixed Effect Specifications

	(1) Ratings	(2) Ratings	(3) Ratings	(4) P("Great")	(5) P(Not "Bad")
Price (pct.)	0.006***	-0.052***	-0.048***	-0.023**	-0.025***
	(0.000)	(0.005)	(0.013)	(0.011)	(0.005)
Pickup	0.035***	0.028***	0.025***	0.020***	0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Late	-0.914***	-0.870***	-0.852***	-0.606***	-0.246***
	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)
Observations	5822058	5822058	5822058	5822058	5822058
Adjusted R^2	0.234	0.317	0.448	0.385	0.440
# of Items	2038040	2038040	2038040	2038040	2038040
Controls	X	X	X	X	X
Week FE		X	X	X	X
Item FE		X	X	X	X
Week X Item FE			X	X	X

Note: This table reports regression coefficients from five separate regressions. An observation is a (rated) transaction item. Outcomes are transaction rating, and the probability that the order (didn't) received the (lowest) highest rating. The independent variable is the natural logarithm transformation, and should be interpreted as percentage changes. All regressions include pickup and delivery dummies, share of item of total order (in monetary terms), whether delivery was late, and day-of-the-week fixed effect. In addition, week, item, and interaction fixed effects are marked below. Standard errors are in parentheses and are clustered at the item-level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2: The Effect of Prices on Ratings Using Sharp Price Changes

	(1) Ratings	(2) Ratings	(3) Ratings	(4) P("Great")	(5) P(Not "Bad")
Price (pct.)	-0.112***	-0.122***	-0.135***	-0.074**	-0.038**
	(0.040)	(0.038)	(0.051)	(0.032)	(0.016)
Pickup	0.029***	0.029***	0.079***	0.022***	0.007**
	(0.010)	(0.009)	(0.014)	(0.008)	(0.003)
Late	-0.853***	-0.864***	-0.751***	-0.607***	-0.246***
	(0.022)	(0.021)	(0.017)	(0.015)	(0.013)
Observations	22512	25004	15460	22512	22512
Adjusted R^2	0.256	0.260	0.108	0.207	0.176
# of Items	6953	7499	6096	6953	6953

Note: This table reports regression coefficients from five separate regressions. An observation is a (rated) transaction item. Outcomes are transaction rating, and the probability that the order (didn't) received the (lowest) highest rating. The independent variable is natural logarithm transformation, and should be interpreted as percentage changes. All regressions include pickup and delivery dummies, share of item of total order (in monetary terms), whether delivery was late, and day-of-the-week fixed effect. Regressions also include an item X price change group fixed effects, where observations are grouped by reviews received around price changes. Standard errors are in parentheses and are clustered at the item-group level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: The Effect of Prices on Ratings for First Time Users Only

	(1) Ratings	(2) Ratings	(3) Ratings	(4) P("Great")	(5) P(Not "Bad")
Price (pct.)	-0.126**	-0.128***	-0.236***	-0.077*	-0.049**
	(0.051)	(0.049)	(0.086)	(0.043)	(0.021)
Pickup	0.035***	0.034***	0.122***	0.026**	0.009**
	(0.013)	(0.012)	(0.021)	(0.012)	(0.004)
Late	-0.862***	-0.876***	-0.674***	-0.595***	-0.267***
	(0.027)	(0.026)	(0.022)	(0.018)	(0.016)
Observations	12472	14258	6477	12472	12472
Adjusted R^2	0.268	0.271	0.093	0.209	0.191
# of Items	3771	4122	2742	3771	3771

Note: This table reports regression coefficients from five separate regressions. An observation is a (rated) transaction item. The sample is restricted only to consumers ordering from the restaurant for the first time. Outcomes are transaction rating, and the probability that the order (didn't) received the (lowest) highest rating. The independent variable is natural logarithm transformation, and should be interpreted as percentage changes. All regressions include pickup and delivery dummies, share of item of total order (in monetary terms), whether delivery was late, and day-of-the-week fixed effect. Regressions also include an item X price change group fixed effects, where observations are grouped by reviews received around price changes. Standard errors are in parentheses and are clustered at the item-group level.
* significant at 10%; ** significant at 5%; *** significant at 1%