Essays on Competition, Digitization, and Innovation

by

Oren Reshef

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven Tadelis, Co-Chair
Professor John Morgan, Co-Chair
Professor Benjamin Handel

Spring 2020

Essays on Competition, Digitization, and Innovation

Abstract

Essays on Competition, Digitization, and Innovation

by

Oren Reshef

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Steven Tadelis, Co-chair

Professor John Morgan, Co-chair

This dissertation comprises three studies on competition, digitization, and innovation. The first study investigates the impact of entry of new competitors on incumbent firms in platform markets, focusing on the heterogeneous effects of firm quality and platform maturity. The second study examines the causal impact of pricing on firm reputation. In particular, online consumer review systems have gained prominence in recent decades, and this study focuses on how these ratings are generated and which factors influence consumer reviews. The third paper focuses on innovation and the impact of individual-level decision-making and characteristics on which patents are issued and the direction of innovation.

In Chapter 1, I study the effect of entry on incumbent firms in platform markets. The entry of firms into a platform has an ambiguous effect on the profitability of incumbent firms operating on the platform: While entry increases competitive pressure on incumbents, supply-side expansion may attract new consumers—effectively increasing total platform size and presumably benefiting all firms. The paper develops a simple model and explores how firm entry affects incumbents' outcomes in a two-sided market. I focus on Yelp Transactions Platform, an online platform that connects consumers with local services. I study a quasi-exogenous increase in firms on the platform and exploit geographic variation to employ a difference-in-differences research design. I find that, on average, market expansion favors incumbents, though the average effect masks substantial heterogeneities: High-quality incumbent firms experience a positive effect, whereas low-quality firms perform unambiguously worse. Using a structural model, my analysis finds a non-monotonic relationship between market expansion and firm performance. Lastly, I use YTP's granular data on consumer and incumbent behavior to explore other market outcomes, main mechanisms, and firms' strategic responses.

The results of Chapter 1 highlight the growing importance of perceived firm quality and standardized ratings on firms' performance. In Chapter 2, I (along with Mike Luca) explore the causal impact of pricing on firm reputation as measured by online ratings. We again use data from YTP on prices, orders, and ratings. Looking at narrow windows around the timing of menu price changes, we find that online reviews are influenced by price changes and that increasing prices tends to harm a firm's reputation; a 1% increase in the price of an

item leads to a decrease of approximately 5% in the ratings left by users. Consistent with this, the distribution of ratings for cheaper restaurants is similar to that of more expensive restaurants. We also find that these effects are not driven by consumer retaliation against price changes, but rather by changes in absolute price levels. Finally, we derive implications to consumers, firms' strategies, and the design of reputation systems.

In Chapter 3, I (along with Abhay Aneja and Gauri Subramani) study how differences in persistence contribute to the gender "innovation gap," i.e. that women are much less likely to receive patents than men. To provide causal evidence of a persistence channel, we use exogenous variation in the likelihood of early-stage adverse decisions about patentability claims that arises from the random assignment of applications to patent examiners. We find that majority-female innovator teams are less likely than majority-male teams to either appeal or amend applications that receive rejections within the patent prosecution process. Roughly 1/2 of the overall gender gap in awarded patents can be accounted for by the differential propensity of women to exit the application process after a rejection of patent claims at the first stage of the prosecution process. We also provide evidence that firms and other organizations can mediate the gender gap in persistence: The persistence gap is reduced for women-led applications that have the backing of firms. We find that examiner identity has little to do with differential persistence across genders.

This dissertation is dedicated to my wife, Eti, for her endless patience and support through good and bad times.

Contents

List of Figures

List of Tables

Acknowledgments

Coauthor information: Chapter 2 of this dissertation is coauthored with Professor Michael Luca of Harvard Business School. Chapter 3 is coauthored with Professor Abhay Aneja from Berkeley Law School and Gauri Subramani of Haas School of Business.

Acknowledgments: I wouldn't have completed my PhD without the help of my dissertation committee chairs. Though they have very different personalities, both of them are among the smartest, sharpest, and most straightforward people I have ever met. First, my academic father, Steve Tadelis, who was always available to support me with a kind word, guidance, a very smart point of view, and a piece of chocolate. He took me under his wing and went above and beyond to push for me in the academic job market. Second, John Morgan has been a tremendous advisor. He made me his apprentice and bestowed upon me his wisdom on what makes a great scholar, teacher, and man. I appreciate and deeply enjoyed the long conversions, in which we often digressed to talk about anything but research.

In addition, Ben Handel, Reed Walker, Noam Yuchtman, Abhishek Nagaraj, and Ned Augenblick provided me with a strong set of research tools and a sense of how to do (good) research. Their support during my PhD studies was invaluable, and I am grateful for their support, feedback, and encouragement. I am also grateful to Mike Luca without whom two chapters of this dissertation wouldn't have been written. This journey was only made possible by Dror Brenner, who pushed me toward graduate studies to begin with.

My time in Berkeley has been wonderful thanks to many friends and colleagues who helped make it my home. They have also been a sounding board for my best and worst ideas, and helped me complete a tricky proof or nail down a regression specification. Special thanks goes to Abhay Aneja, Dayin Zhang, Jean Zeng, Troup Howard, Gauri Subramani, Charu Gupta, Petr Martynov, Sibo Lu, and Laura Boudreau. I hope to continue to develop our research relationships and friendships in the years to come

Those outside the research community who helped me reach this point deserve an even greater acknowledgment. My parents, Maya and Asher, were always more than ready to put their lives on hold in order to help and support us. We wouldn't have made it through without them. The same goes to my parents-in-law, Larisa and Soso. My friends, especially Ran and Onn, that kept me sane and smiling all of these years. Most important of all is my wife, Eti, who came with me halfway across the world just to make my dreams come true and has been with me every step of the way. Finally, little Leah and Ethan reminded me there are more important things in life than research.

Chapter 1

Smaller Slices of a Growing Pie: The Effects of Entry in Platform Markets

Chapter abstract: The entry of firms into a platform has an ambiguous effect on the profitability of incumbent firms' operating on the platform: While entry increases competitive pressure on incumbents, supply-side expansion may attract new consumers—effectively increasing total platform size and presumably benefiting all firms. Guided by a simple model, this paper explores how firm entry affects incumbents' outcomes in a two-sided market. Specifically, I focus on Yelp Transactions Platform, an online platform that connects consumers with local services. I study a quasi-exogenous increase in firms on the platform and exploit geographic variation to employ a difference-in-differences research design. I find that, on average, market expansion favors incumbents, though the average effect masks substantial heterogeneities: High-quality incumbent firms experience a positive effect, whereas low-quality firms perform unambiguously worse. Using a structural model, my analysis finds a non-monotonic relationship between market expansion and firm performance. Lastly, I use YTP's granular data on consumer and incumbent behavior to explore other market outcomes, main mechanisms, and firms' strategic responses.

1.1 Introduction

Platform markets have now spread over diverse sectors of the economy, including retail (eBay, Amazon, Taobao), travel (Airbnb, Homeaway), services (TaskRabbit, Upwork, Uber), and finance (Kickstarter, Ant Financial).[1] As platforms spread to diverse areas of the economy, more firms and sellers begin to operate within a platform settings. It is thus important to understand the impact of platform growth and firm entry on the incumbent firms operating on the platform and platform dynamics. In traditional markets, entry of new firms exacerbates competitive pressure on incumbents and erodes their excess profits (Porter, 1989, 1997, or textbook analysis in Samuelson, 1951). However, in two-sided markets, entrants

[1]In fact, out of the ten firms with highest market capitalization, seven are platforms: Apple, Amazon, Alphabet, Microsoft, Facebook, Alibaba, and Tencent. Moreover, of the 2018 most promising 'unicorns'—start-ups with valuation of over $1 billion—about 60 to 70 percent were platform (Cusumano et al., 2019).

may have an indirect positive effect on incumbent firms because entrants can increase the appeal of the platform to potential consumers. If this sufficiently increases total demand, then market expansion can more than compensate for competitive pressure.

In this paper, I consider the net effect of these competing forces within a major platform market. Specifically, I address three related questions: First, how does new-firm entry affect the performance of incumbent firms in platform markets? Second, which types of incumbent firms benefit or lose the most from entry? And third, do incumbents readjust their strategies to respond to entry, and if so, how? Guided by a stylized model, I answer these questions using proprietary data from the Yelp Transactions Platform (YTP), an online platform for takeout and delivery from local restaurants. I study YTP's partnership with Grubhub delivery service, which substantially and suddenly increased the number of sellers on the platform without directly affecting the number of consumers.

I use the sharp change and geographic variation in the effect of the partnership to employ a difference-in-differences research design and find that entry benefits high-quality incumbents and raises their revenue by up to 15.8%. In contrast, entry hurts low-quality incumbents and reduces their revenue by as much as 9.2%. Hence, entry in two-sided markets changes the nature of competition by increasing the returns to higher quality and the average quality purchased. These results have important implications for firms' platform incentives and strategies, for platform designers, and for competition policy. I complement the reduced-form results with a structural model of consumer participation and product selection, which allows me to extrapolate the results to a broader set of market conditions.

The paper fills a substantial gap in knowledge because empirically studying entry, especially in two-sided markets, is challenging given that entry and competition are often hard to isolate and measure. One characteristic muddling empirical analysis is the fact that the relevant set of competitors depends heavily on the market's definition, which may vary over time. Additionally, entrants make strategic decisions about where and when to enter, and each decision will correspond to the dynamics within and outside the immediate market. As such, entry into markets is endogenous and likely correlated with unobserved market characteristics or shocks that confound estimating the impact of entry. Consequently, while there is a huge theoretical literature on indirect network externalities and two-sided markets (Katz and Shapiro, 1985, Rochet and Tirole, 2006), the empirical literature is sparse.

I overcome these challenges using proprietary data from Yelp—the commonly used consumer-review website for local businesses—as well as YTP purchases to study all food ordering- and delivery-transactions finalized on YTP over a period of almost two years. Thanks to the type of services provided and the granular data that Yelp collects, I am able to clearly define the relevant market, market participants, and outcomes of interest. The core sample includes almost 50,000 incumbent establishments in almost 4,000 municipalities, and the main outcomes of interest are establishment-level weekly ordering quantities and weekly revenue from the platform. I also take advantage of Yelp's core data, which includes business attributes and information on how consumers perceive establishments. Subsequent analysis uses granular data on item-level pricing, firms' advertising expenditures on Yelp, and consumers' search and review behavior.

Using these data, my empirical strategy exploits the event of YTP's partnership with Grubhub, the largest food delivery service in the United States. Notably, YTP only *connects* users with delivery services (partners); these partners establish agreements with individual restaurants and handle the deliveries themselves. Thus, the number of restaurants on the platform depends on YTP's delivery partners. Following the Grubhub partnership that commenced in February 2018, YTP users instantly gained access to Grubhub's extensive network of affiliated restaurants, sharply increasing the number of businesses available on YTP. While the overall change in the number of businesses was substantial, there was significant variation in the impact of the partnership across geographical regions (For example, the number of restaurants available on the platform significantly increased in Berkeley, California, but remained largely unchanged in the neighboring city of Richmond). I use quasi-experimental variation in the intensity across geographic regions to employ a difference-in-differences research design in order to study the causal effect of market expansion through firm entry on the performance of incumbent businesses. Since Grubhub's network of affiliated restaurant is not randomly assigned, I perform several tests to verify the validity of the research design, as described in Section 1.4.

To guide my empirical analysis, I develop a stylized model to demonstrate the countervailing effects of entry on incumbents' performance. Initially, holding fixed the number of users on the platform, entrants compete with incumbents to capture participating consumers, and reduce the incumbents' market share. On the other hand, consumers choose whether they wish to join the platform based on their beliefs about how attractive the platform is. When more firms are present in the market, consumers' expected benefit from joining the platform increases, leading to higher participation, which increases the total market size. Increased overall demand may in turn benefit all firms in the market, offsetting the negative effects of the reduced market share on incumbents' profits. Simply put, when more firms enter, incumbents receive a smaller slice of a larger pie. Thus, whether incumbents benefit on net is an empirical question. The model in turn produces several testable predictions, most notably that higher quality firms will benefit more (or suffer less) from entry.

I find that the entry of new firms *increased* incumbent firms' weekly revenue on average by 4.5% in treated markets as compared to untreated markets. This result is best explained by increases in total platform demand: markets that experienced an increase in the number of restaurants were able to attract 36% more consumers to the platform and increase platform-level revenue by almost 60%, as compared to untreated markets. Notably, while the positive average effect is consistent with the hypothesis of countervailing forces affecting firms, the effect masks considerable heterogeneity in firms' outcomes. In particular, guided by predictions from the model, my analysis tests for heterogeneous effects across firm quality.[2] I find that entry *increased* weekly revenue of high-quality incumbent firms by 9.8%–15.8%. In contrast, entry *decreased* weekly revenue of low-quality firms by up to 9.2%. These results suggest that the total effect depends on firms' characteristics: Entry changes the nature of competition on the platform, increases returns to high ratings, and improves the average quality purchased by consumers. Additionally, I find that differentiation helps mitigate the negative impact of entry: Incumbents that were similar to entrants suffered a 3%–6% larger revenue loss.

In the second part of the analysis, I examine firms' responses to changes in their competitive environment and the forces driving the main result. Platform expansion intensified the positive relationship between firm quality and performance. Accordingly, I find suggestive

[2]Quality is defined by the relative Yelp star rating in the market on the eve of integration (Section 1.4.1).

evidence that firms responded to this change in incentives and increased their subsequent investments in quality: The average quality of firms in treated cities rose by about 1% more than control cities, an increase associated with a 3% growth in revenue. I find that entry also affected firms' advertising behaviors on Yelp. In particular, average advertising *decreased* in treated markets, and this effect was driven solely by highly-rated establishments. This finding suggests that market size and advertising operate as substitutes, because a firm's improved performance reduces the benefits from advertising. Finally, while the partnership positively affected revenue on YTP, the effect on total firm revenue is unclear. One possibility is that new consumers may be substituting away from other platforms, leaving total revenue unchanged; however, using Yelp's search data, I provide suggestive evidence that appears inconsistent with strong substitution patterns across platforms. I am also able to rule out that the main effects are driven by the ordering of search results.

I then turn to study the impacts of initial platform size. In the case analyzed here, YTP had (on average) only 4.2% of a city's total number of restaurants even after Yelp and GrubHub partnered. It thus remains unclear how the main results extend to settings where total market size is larger—for instance, will the entry of new firms continue to increase high-quality incumbents' revenue when 80% of a city's restaurants are already present on the platform? To tackle this question, I impose additional functional-form assumptions on the theoretical model. I then estimate a structural model of discrete choice and consumer entry to characterize consumers' utility function and entry-cost distributions. Using these structural estimates, I simulate firm performance for the full range of market saturation, up to the point where all firms participate on the platform.

The simulations yield several results. First, entry improves consumer welfare by 32% for the median affected market; welfare continues to increase as the fraction of establishments on the platforms grows, but at a decreasing rate. Second, high-rated businesses perform better than low-rated businesses at all levels of market entry. Finally, for almost all firms (i.e., excluding those with extremely low ratings), there is a non-monotonic, inverted U-shape relationship between the percentage of businesses on the platform and sales: When the fraction of firms on the platform is relatively small, the *market size* effect dominates and sales increase as more competitors enter the market. In contrast, when a sufficient number of businesses enter the platform, the *market share* effect dominates, and sales drop with each additional entry. Moreover, there is a monotonic positive relationship between firm rating and the number of firms on the platform needed to maximize sales (i.e., higher-rated firms prefer higher levels of firm participation on the platform). For instance, the main estimates suggest that the bliss point, the point at which profits are maximized, is about 46% participation for top-rated firms and only 11% for the median-rated firm.

This paper contributes to several strands of the literature. First, this work builds on the theory of network externalities and two-sided markets, as introduced by Katz and Shapiro (1985) and Farrell and Saloner (1985). Identifying indirect network externalities is extremely difficult and is usually estimated from just one market, by comparing two competing technologies, or by imposing structural assumptions on network development. (See, e.g.,Rysman, 2004, Springel, 2018, Ohashi, 2003, Corts and Lederman, 2009, Stremersch et al., 2007, Lee, 2013 and Nair et al., 2004). I add to this literature by providing clear evidence identifying the importance and magnitude of indirect network externalities across many markets.

Second, this paper focuses on the performance of the firms operating on the platform, and offers new insights on the forces effecting platform participants. Previously, the two-sided markets and platform-strategy literatures predominantly emphasized platform-level outcomes and strategy, such as platform pricing, platform compatibility, competition between platforms, and efficient platform size.[3] More recently, a small literature has begun to explore the positive spillovers generated by entry in two-sided markets (see, e.g., Li and Agarwal, 2016, Cennamo et al., 2016, Mahajan et al., 1993, Shen and Xiao, 2014, Cao et al., 2018), with most work focusing on the spillovers generated across platforms. I extend this literature by studying the expansion of sellers *within* a platform; I offer a new mechanism to generate positive spillovers across competitors and examine heterogeneous effects by firm quality and platform maturity, which remains mostly unexplored in the previous literature.

Third, and more generally, this work contributes to the industrial organization literature on entry and its effect on incumbent firms. Starting with the seminal paper of Bresnahan and Reiss (1991), the empirical literature has mostly focused on the competitive effect of entry and the downward pressure on prices (see, e.g., Berry, 1992, Nickell, 1996, Syverson, 2004, Hortaçsu and Syverson, 2007, Jia, 2008) and product variety (Illanes and Moshary, 2018). Additionally, this work has normative implications to the analysis of barriers to entry (see, e.g., Goolsbee and Syverson, 2008, Hauser and Shugan, 2008, Seamans, 2013, Kadiyali, 1996, Fudenberg and Tirole, 1984, Ellison and Ellison, 2011). In particular, I argue that the profitability of barriers depends on the characteristics of firms and that, in certain cases—such as two-sided markets—firms can perform better by welcoming entry rather than deterring it.

Finally, following the seminal work of Melitz (2003), the trade literature studied the effects of entry into new foreign markets and, similar to this paper, finds that entry differentially impacts incumbents as a function of their quality (e.g., Aghion et al., 2009, Pavcnik, 2002). Though my work focuses on consumers' information frictions as the source of positive externalities, the main results apply to alternative mechanisms posited in prior literatures, which suggests that the forces described in this paper may be extended to other settings. Some prominent examples include (1) agglomeration effects (e.g., Marshall, 1890, Murphy et al., 1989); (2) new product discovery (e.g., Bass, 1969, Fosfuri and Giarratana, 2009); and (3) taste for variety (Dixit and Stiglitz, 1977).This paper also contributes to our understanding of online-rating mechanisms and how they develop as platforms grow (Luca, 2016, Chevalier and Mayzlin, 2006, and the review in Tadelis, 2016).

The remainder of the paper is organized as follows. Section 1.2 presents the conceptual framework that motivates the empirical analysis. Section 1.3 gives an overview of YTP and its partnership with Grubhub. Section 1.4 lays out the data and empirical strategy. Section 1.5 presents the results. Section 1.6 describes the structural model and simulation results. Section 1.7 provides conclusions and outlines new opportunities for future research.

[3] Some prominent theoretical and empirical analyses at the platform level include: Jin and Rysman (2015), Seamans and Zhu (2013), Kaiser and Wright (2006), Cullen and Farronato (2019), Zhu and Iansiti (2012), Farronato and Fradkin (2018), Gawer and Henderson (2007), Zhu and Liu (2018), Dubé et al. (2010), Caillaud and Jullien (2003), Rochet and Tirole (2006), Armstrong (2006), and Hagiu and Wright (2015).

1.2 Conceptual Framework

The conceptual framework develops a model of consumer choice and its effect on firms' outcomes. This theoretical analysis guides the empirical analysis and is the basis for the structural estimation presented in Section 1.6.

1.2.1 Setup

The market consists of a unit continuum of consumers. Each consumer is interested in purchasing exactly one unit of the product or service and may join the platform to shop for a product. Consumers who do not join the platform receive an outside utility ω, which is the same for all consumers.[4]

The platform has a menu of N firms from which consumers may choose. Firms are characterized by quality, which can take two values, $q_j \in \{0, q\}$ with equal probability, where $q > 0$. Purchasing a product generates a mean utility equal to the seller's quality, q_j. Thus, q captures vertical differentiation between sellers. In addition, firms are horizontally differentiated in the sense that consumer i buying from seller j also receives an idiosyncratic utility shock, $\epsilon_{ij} \sim G(\epsilon)$. The utility consumer i gains from buying from seller j is given by $U_i = q_j - p_j + \epsilon_{ij}$. I denote the distribution of the difference between two random variables distributed $G(\cdot)$ as $\tilde{G}(\cdot)$

In order to use the platform, consumers must pay a one-time user-specific fixed cost, $c_i \sim H(c)$. This cost may be either monetary or the hassle cost associated with setting up an account and learning how to use the platform efficiently. Users are fully rational and know the fundamentals of the model as well as their individual entry cost,c_i. They form (correct) beliefs on the number of firms available on the platform, their prices, the number of high and low quality firms, and the distribution of entry costs and idiosyncratic shocks. They do not, however, know the realization of the idiosyncratic utility shock, ϵ_{ij} or the prices on the platform.[5] The search engine of the platform is extremely efficient and, conditional on using the platform, consumers immediately and costlessly observe all the sellers available on platform as well as the idiosyncratic utility shock. Consumers always choose the seller (or the outside option) that maximizes their utility.

Firms know their quality, q_j, and the quality of all other firms in the market,q_{-j}, but do not know the idiosyncratic shock, ϵ_{ij}. Firms face a constant marginal cost of r to provide one unit of the product. They compete in Bertrand-Nash competition and set prices to maximize profits.

The timing of the model is as follows: First, firms form beliefs about other firms' strategies and consumers' entry decisions, and consumers form beliefs about firms' prices and the expected value from joining the platform. Second, firms set prices. Third, entry costs are realized and consumers choose whether to join the platform. Finally, random utility shocks are realized and consumers choose the best option available to them.

[4]All of the results hold with heterogeneous ω_i. I impose this restriction for the structural analysis primarily because it is difficult to separately identify heterogeneous outside options, ω_i from the heterogeneous entry costs, c_i (presented below). The structural analysis, however, allows ω to differ across *markets*.

[5]Unobserved prices are consistent with the classical search literature (e.g., Stigler, 1961, Burdett and Judd, 1983).

1.2.2 Analysis

Given that all firms of the same quality level a priori face the same demand and marginal costs, in a symmetric equilibrium, $p_j = p_h \quad \forall j \in N_h$ and $p_k = p_l \quad \forall k \in N_l$, where N_h and N_l are the sets of high- and low-quality firms, respectively.

Lemma 1 *Under some restrictions on $G(\cdot)$, high-quality firms charge higher prices but provide greater mean utility to consumers than low-quality firms, $p_h > p_l > p_h - q$.*

Proof All proofs are in Appendix A.2.

After joining the platform, consumers enjoy zero search costs and full information. Thus, they always choose the firm j that maximizes their utility, $j = \arg\max u_{ij}$. Consumers, however, will join the platform only if their (expected) gain from joining outweighs their outside option plus the realized entry cost: $E_\epsilon[max(u_{ij})] > \omega + c_i$.

Given consumers' behavior, firms' profit functions for a high- and low-quality firm j can be written, respectively, as:

$$\begin{aligned}\Pi_j^h = &H(E_\epsilon[max(u_{ij})] - \omega) * (p_j - r) * \\ &P(\{\epsilon_{ij} > \epsilon_{ik} - p_h + p_j, \forall k \in N_h\} \cap \{\epsilon_{ij} > \epsilon_{im} - p_l + p_j - q, \forall m \in N_l\} \\ &\cap\{\epsilon_{ij} > \omega + p_j - q\})\end{aligned} \tag{1.1}$$

$$\begin{aligned}\Pi_j^l = &H(E_\epsilon[max(u_{ij})] - \omega) * (p_j - r) * \\ &P(\{\epsilon_{ij} > \epsilon_{ik} - p_l + p_j, \forall k \in N_l\} \cap \{\epsilon_{ij} > \epsilon_{im} - p_h + p_j + q, \forall m \in N_h\} \\ &\cap\{\epsilon_{ij} > \omega + p_j\})\end{aligned} \tag{1.2}$$

The profit functions have an intuitive interpretation: The first element captures the share of users actively searching on the platform and corresponds to total *market size*. This element has the same magnitude for high- and low-quality firms. The second element is the per-unit markup. By Lemma 1, this second element is strictly larger for high-quality firms. The third element represents the probability that users choose to purchase from firm j. In particular, firm j is chosen by consumer i if the firm generates higher utility than all high-quality firms, low-quality firms, and the outside option (the entry cost is sunk). This term corresponds to the firm's *market share* of the users actively searching on the platform. This term is strictly larger for high-quality firms, since they face less competition from both high- and low-quality firms. These intuitions are formalized in the following lemma:

Lemma 2 *In equilibrium, high-quality firms sell more and generate higher revenues and profits than low-quality firms, $\Pi_j^h > \Pi_j^l$.*

The main goal of the analysis is to study the impact that entry of new sellers has on incumbent firms' profits. The following proposition motivates the two main specifications of the empirical analysis:

Proposition 1 *The effect of entry on firms' profits and revenue is:*
(1) Ambiguous, $\frac{\partial \Pi^h}{\partial N}$ and $\frac{\partial \Pi^l}{\partial N}$ can be positive, negative, or zero.
(2) More positive (or less negative) for high-quality firms compared to low-quality firms, $\frac{\partial \Pi^h}{\partial N} > \frac{\partial \Pi^l}{\partial N}$.

The intuition behind the first part of Proposition 1 is straightforward: On the one hand, increasing the number of firms in the market will positively affect market size; as the number of firms increases, the expected value of the *best* product increases as well, drawing more consumers into the market. On the other hand, as the number of firms increases, the probability of any specific firm to provide the highest value decreases, eroding its market share. The relative magnitudes of these two forces determine the total impact of entry on incumbent firms.

The second part of Proposition 1 stems from the fact that both high- and low-quality firms benefit from the increase in total market size, but there is an inherent asymmetry in the effect of entry on market share. To see this, consider the case where q is extremely large. In that case, adding many low-quality firms will hardly affect the market share of a high-quality firm. Inversely, adding just one high-quality firm is likely to substantially reduce the market share of the low-quality firms. The same intuition holds in general: low-quality firms hurt more from new competitors joining the market compared to high-quality firms.

1.3 Setting

I apply this framework to a portion of the food delivery–service industry in the United States, a $35 billion dollar industry that is expected to grow at an average annual rate of more than 20% in the next 10 years.[6] The main focus is the Yelp Transactions Platform, an online platform launched in 2013 by the consumers' review website, Yelp. YTP enables users to order food delivery and pickup from local restaurants through several food-delivery services.[7] The empirical analysis focuses on YTP as the relevant market in which restaurants compete for users interested in food-delivery services.

YTP operates as a part of the standard Yelp website and features a subset of restaurants available on Yelp. Shoppers are automatically directed to the platform by applying the "Delivery" or "Takeout" filters, or by using similar words in a search query.[8] Figure 1a depicts the results of a search query on YTP: the shopper views a list of restaurants relevant to the query and user's location as well as a map of the establishments. Restaurants' data are pulled from the standard Yelp website and include the star rating, number of reviews, food category, and 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 Figure 1b. Consumers can then choose the specific menu items they are interested in and finalize the transaction. Notably, the entire order process is native to YTP and is finalized without users being redirected to external websites.

Importantly, YTP is not a delivery service but a marketplace where consumers can find food delivery and specific restaurants; delivery and takeout transactions must be implemented and handled by third-party delivery services, referred to as "partners." For example,

[6]According to a UBS Investment Bank report.

[7]YTP allows users to transact with a myriad of other local businesses, including hotels, home services, and local services (such as doctors and mechanics). Due to the nature of the institutional shock of interest, this paper restricts attention to food ordering.

[8]There are alternative ways to access YTP: First, shoppers can start an order directly from the business page. Second, when a user performs a search on the standard Yelp website, businesses that are YTP affiliates will have an "Order Now" button next to their name on the search results.

as seen in the bottom-right corner of Figure 1b—the order is carried out by Grubhub. Accordingly, Grubhub processes the order and sends the details to the restaurant; in the case of delivery, a Grubhub employee will also pick up the prepared food and deliver the meal to the customer's address. Other delivery services partnering with YTP include Delivery.com, ChowNow, Eatstreet and more. The identity of the delivery service is determined automatically by the YTP algorithm and cannot be changed by the user. YTP partners with delivery services and *not* with specific restaurants. Thus, in order for a restaurant to appear on YTP, the restaurant must first contract with a delivery service, which can, in turn, sign a partnership agreement with YTP. Thus, the supply of restaurants on YTP is determined by its partners and their network of affiliated restaurants.

In August 2017, Yelp entered into a new partnership with Grubhub (the largest food delivery service in the United States to date) that would allow users to begin ordering from Grubhub through YTP beginning in February 2018.[9] In return, Grubhub agreed to pay Yelp a fixed fee for every order sent through YTP. Thanks to this partnership, YTP users were able to access Grubhub's extensive network of restaurants, nearly doubling the number of restaurants on YTP. Grubhub, on the other hand, gained access to Yelp's enormous user base.[10] [11]

The partnership between YTP and Grubhub launched mid-February 2018, with the systems gradually integrating across platforms (iOS, Android, www) and geographical areas thereafter. Integration was (formally) finalized on March 19, 2018. The research design exploits the sharp and dramatic increase in the number of restaurants available on YTP in order to analyze the impact on existing restaurants.[12]

1.4 Methodology and Research Design

1.4.1 Data

This section provides an overview of the data used in the paper. I focus on the core pieces of data required for my findings here and relegate a more detailed account to Appendix A.3. Table 1 presents the descriptive statistics.

[9]As part of the agreement, Yelp sold its subsidiary delivery service, Eat24, to Grubhub. Yelp acquired the service in 2015 for almost half the sale price. While the service had substantial presence in several areas, such as San Francisco and Miami, its market share and network of affiliated restaurants were both relatively small compared to Grubhub's network. This agreement essentially meant that Yelp was exiting the operational side of delivery and focusing on the online interface with consumers.

[10]The strategic alliance was based on the notion that Grubhub would supply the restaurants and Yelp would supply the buyers, as is evident in the joint press release following implementation: " [...] combination of Grubhub's unmatched restaurant network and efficient delivery infrastructure with Yelp's large purchase-oriented audience. [...] Yelp users will be able to order from far more local restaurants [...] from Grubhub's huge network of local favorites."

[11]Note that the partnership did not imply exclusivity: Grubhub restaurants remained active on its own website, and YTP still featured businesses affiliated with other delivery services.

[12]To be conservative, I code February 19, a month before completing the integration, as the first week of treatment. In Section A.4.2 I discuss alternative definitions of the integration dates: First, defining the integration date as March the 19th, and second, excluding the period between February to March 19th. In general, the main estimates remain statistically significant and are larger in magnitude when using either of these alternative definitions.

I use proprietary Yelp and YTP data covering a period of almost two years, from the beginning of 2017 to the end of 2018, with the platform integration occurring approximately at the middle of the period—beginning in February 2018. The data include all food orders completed on YTP during that period.[13] For each transaction, I observe item-level description and price, the date on which the order was made, the identity of the user and business, and the delivery partner. I aggregate transactions to the business-week level; the main outcomes of interest are the number of weekly orders and total weekly revenue, excluding tips, taxes, and delivery fees. The final sample includes 56,493 establishments and over four million business-week observations.

The data represent all restaurants in cities where YTP is available. My business data include the Yelp business ratings, the dates on which the business joined and exited YTP, the type of food sold, the business address, and the platform's Dollar Ratings.[14] YTP ratings are based on Yelp's Star Rating system, a user-generated rating on a one- to five-star scale. While the Yelp ratings presented to users are rounded to the nearest half star, in my analysis I use the underlying, continuous, rating. High- and low-quality businesses are defined based on their Yelp rating on the eve of integration, as opposed to the rating on a specific week, since subsequent ratings might react to treatment assignment. I define high- and low-ratings in two alternative ways: (1)*binary definition*: indicator for above- or below-median rating in the relevant geographical area; and (2)*sharp binary definition*: indicator for above the 75th percentile or below the 25th percentile rating in the relevant geographical area, i.e., above or below the median rating, excluding the interquartile range. It is valuable to note that both of these definitions also include new entrants when calculating the rating percentile, which I have chosen to do to control for the differences in quality of entering firms across markets. I also collect similar data on restaurants that do not participate in YTP; I use these data to conduct several placebo tests. Descriptive statistics on firms' ratings, dollar ratings, and tenure on YTP are presented in Panel A for Table 1.

The data include documentation of consumers' search processes. In particular, I use data on total usage of YTP and the number of users interested in delivery on the platform (see Section 1.6 for details). I also take advantage of the characteristics of search sessions that ended in a food order, including the number of searches, the number of business views, the session duration, and the search queries. Finally, I collect data on businesses' daily advertising expenditures on Yelp.

For the structural estimation in Section 1.6, I use several external sources to collect city demographics: County-level data on total population, gender, age, and income come from the American Community Survey 5-Year Data for 2017. I use the IRS Individual Income Tax ZIP Code Data (2016) to provide zip code-level data on annual gross income. These zip code-level data are joined with business location to approximate the type of neighborhood the business occupies (downtown, suburbs, etc.).

[13]Unfortunately, as part of my agreement with Yelp, I am unable to disclose sensitive business information regarding the *levels* of platform or business performance. I cannot disclose, for instance, the total number of orders or users on the platform, the number of orders per business, revenue, or the precise number of users. Accordingly, all of the main results will be presented as a percentage change rather than absolute values.

[14]Dollar Ratings are meant to approximate the overall cost per dinner, and are assigned by users and aggregated by Yelp. Dollar ratings on Yelp take on four discrete values: $ = under $10, $$=11-30, $$$=31-60, and $$$$= over $61.

Market Definitions Since fresh food can only be delivered within a reasonable distance, in this paper's context, markets are naturally defined by geographical areas. Accordingly, in the main analysis, I define a market by a city-state combination (3,965 markets). There are several reasons to choose this definition: First, data are most complete for restaurants' and users' city, as opposed to county or zip-code. Second, users search on the platform by city when submitting a query. Third, delivery areas are often bounded by arbitrary limits as opposed to real distance. Finally, though I do see some deliveries across city limits, the vast majority of deliveries takes place within a given city.[15]

While such market definitions are logical, a potentially appealing alternative is to use city-food category combinations as the relevant market, i.e., San Francisco pizzerias will be considered a different market than sushi restaurants in San Francisco. I discuss these alternative market definitions as part of the robustness checks in Section 1.5.3.

Treatment Intensity Treatment intensity is defined as the change in percentage of restaurants on YTP out of the total restaurants in the city following the Grubhub integration. I approximate the total number of restaurants in the city by using the total number of restaurants featured on Yelp. Formally, treatment intensity is defined as:

$$TI = \frac{\text{\# restaurants on YTP Post - \# restaurants on YTP Pre}}{\text{Total numbers of restaurants on Yelp}} \tag{1.3}$$

To give a concrete example, if a city has a total of 50 restaurants, and 20 were listed on YTP before the partnership with Grubhub, when 10 new businesses were added following the partnership, the treatment intensity would be coded as 20% (10 over 50). Thus, treatment intensity captures the *change in percentage points* of restaurants on YTP. It is important to standardize the absolute number of restaurants added by potential market size since we would expect an addition of 100 new restaurants to have different implications in very large metropolitan areas compared to small towns.

Table 1 presents descriptive statistics on YTP participation and the average treatment intensity. Prior to the partnership with Grubhub, only a small percentage (3.2% on average) of restaurants in a city were available on YTP. While the increase in the share of businesses on the platform is substantial relative to the baseline, even after integration, on average, only 4%–5% of restaurants in the city are available on the platform.[16] [17]

[15]Nevertheless, the main results are robust to alternative definitions of local markets, including county, three-digit zip code, and five-digit zip code.

[16]Notably, the average treatment intensity at the business level is substantially larger than at the city level, 2% and 1.2%, respectively. The reason for this difference is that treated cities tend to be larger and have more businesses, thus pulling the mean upwards. I address this issue in Section 1.5.3, Robustness Checks.

[17]An alternative definition of treatment intensity uses the *percentage change* in the share of restaurants on the YTP, e.g., in the example above, treatment intensity will be coded as 50% (10 over 20). The main issue with this definition is that it mechanically introduces very large intensities in small places. In an extreme example from the data, a city with only two businesses prior to integration receives a treatment intensity of 600%. Section 1.5.3 shows that the main results are robust to the alternative definition of treatment intensity.

1.4.2 Research Design

The partnership between YTP and Grubhub provides a platform-level institutional shock to the number of restaurants on the platform. Figure 2 presents the change in the number of businesses over time. As in all following figures, the first week of treatment, February 19, 2018 (8th week of 2018) is normalized to nearly zero. We can see a substantial and discontinuous increase in the number of restaurants on the platform in a short period of time; the number of restaurants rises by over 60% in just a few weeks.[18]

While the aggregated effect on the number of restaurants available on YTP is substantial, there is significant variation in the impact across cities. In fact, the median number of businesses added to a city following the Grubhub integration is zero. Figure 3 presents the distribution of treatment intensities. Figure 3a presents the distribution of the change in the percentage of restaurants available on YTP in a city. A little less than half of the cities were not affected by the partnership, and the vast majority of cities experienced an increase of less than 5% in the percentage of firms available on the platform. Figure 3b displays the distributions of the percent of restaurants on the platform out of the total number of restaurants in the city before and after the YTP-Grubhub partnership, conditional on a non-zero change.

I use the regional variation in treatment intensity to employ a difference-in-differences analysis, wherein I compare cities with little or no change in the share of restaurants available on YTP to cities with larger changes. The key identifying assumption is that the treated cities would have had similar trends to the control cities in the absence of the Grubhub integration. Though the parallel trend assumption is not directly testable, I offer several pieces of evidence to suggest that it holds in my setting. Appendix A.4.1 presents a detailed discussion of the robustness checks as well as the results for all of the tests I performed, as summarized in the following paragraphs.

First of all, variation in entry originates from a platform-level institutional change, which is unlikely to be correlated with unobserved city-level trends. Second, the two main outcomes of interest, weekly revenue and number of orders, for businesses in treated and control cities trend similarly in the period preceding the partnership between Yelp and Grubhub (see Section 1.5), which is consistent with the parallel trends assumption. Moreover, I conduct a placebo test, in which I counterfactually set the integration date to the middle of the pre-treatment period; I do not find any differences in trends in the pre-partnership period when looking at all business, nor do I find differences when restricting attention only to high- or low-rated firms separately, which are the three main specifications used in the empirical analysis (results are presented in Appendix Table A1).

Third, if treatment effects are driven by some other, unobserved, shifts in trends that are unrelated to the Grubhub partnership, then we can expect to find significant differences in other city-level outcomes not directly related to food ordering. I conduct several placebo tests to examine whether the partnership is correlated with outcomes of *non-YTP* businesses, such as the number of businesses on Yelp, the average rating, and the number of reviews per business. I find null effects on all of these dimensions (Appendix Table A2). Finally, in this

[18] I attribute all new businesses added in the first eight weeks following week zero to the Grubhub integration. Over 98% of these businesses have Grubhub as their first delivery partner. I do not include businesses added after week 8, since later additions might be the result of different trajectories of market development.

setting, treatment intensity is determined by the presence of YTP and Grubhub in a given city and the overlap between their networks of restaurants in that city, which are randomly assigned. I find that treatment assignment is correlated with city characteristics: Treated cities are, on average, larger and have more restaurants and a higher share of restaurants on YTP. To address potential concerns that initial difference are driving the main results, I test several alternative specifications: (1) Inverse probability weighting, which accounts for the different probabilities of selection into treatment based on observables. (2) A more demanding analysis in which I reassign treatment by propensity score bins. The main results of the analysis are robust to all of these specifications (Appendix Table A6).

1.4.3 Empirical Specification

The primary empirical specification takes the difference-in-differences functional form:

$$Y_{jt} = \beta Post_t * Treat_j + \gamma_j + \delta_{st} + \epsilon_{jt} \tag{1.4}$$

Where t is the index for the week, j denotes the unit of observation (establishments in the main specification, and city when analyzing aggregate effects), and s is an index for the state. Y denotes the outcome of interest: weekly revenue and the number of orders, in the main specification. $Post$ is a binary indicator variable for whether the partnership came into effect. $Treat$ captures the treatment intensity of unit j and takes three forms: (1) *Binary treatment*, an indicator that signifies whether the treatment intensity in the city is above-median intensity across all cities. (2) *Sharp binary treatment*, a binary indicator variable that takes the value 1 whenever treatment intensity in that city is above the 75th percentile and 0 whenever treatment intensity is below the 25th percentile; since the median treatment intensity is almost zero, this sharp binary treatment definition effectively compares cities with no change to cities with treatment intensities above the 75th percentile.[19] (3) The continuous measure of treatment intensity, as defined in equation 1.3; this quantity represents the (continuous) change in the percentage of firms on YTP out of the total number of firms in the city. γ_j denotes unit-level (business or city) fixed effect, and δ_{st} denotes state-week fixed effects.[20] Standard errors on all regressions are clustered at the city level, which is the level at which the treatment is administered. The parameter of interest, β, captures the causal impact of increasing the share of businesses on the outcome variable.

Motivated by the model, subsequent analyses estimate heterogeneous treatment effects by firms' quality, as measured by Yelp business ratings, using the difference-in-difference-in-differences framework given by:

$$\begin{aligned} Y_{jt} = {} & \beta_1 Post_t * Treat_j + \beta_2 Post_t * Treat_b * Low_j \\ & + \beta_3 Post_t * Low_j + \gamma_j + \delta_{st} + \epsilon_{jt} \end{aligned} \tag{1.5}$$

[19]The rationale behind the last definition is to omit cities that are only weakly treated, which enables this research design to compare cities that received no treatment with cities that experienced meaningful treatment intensity. I use this last treatment definition in most figures and tables.

[20]The regression equation does not include any of the business-level covariates described in Section 1.4.1, because, apart from Yelp business ratings, all covariates are constant over time and are absorbed by the unit fixed effect. Ratings are not included in the regression since these are potentially affected by treatment assignment (see Section 1.5.4), which may consequently bias the estimates.

Here—beyond the indices and variables defined in equation 1.4— Low_j serves as the binary variable defined in Section 1.4.1, which captures the relative ranking of businesses at the eve of integration. β_1 captures the impact of entry on outcome Y for highly-rated businesses. Similarly, $\beta_1 + \beta_2$ captures the causal impact of entry on outcome Y for low-rated businesses. Intuitively, the heterogeneous impacts are estimated from the differential changes between treated and control cities for the relevant subset of businesses, i.e., comparing high-quality firms in treated markets to high-quality firms in untreated markets, and similarly for low-quality firms.

1.5 Results

1.5.1 Aggregate Outcomes

I begin by assessing the impact of the supply-side expansion on total market size. Figure 4 depicts the percent differences in weekly aggregated market-level outcomes between markets that are above the 75th percentile and below the 25th percentile of treatment intensity (*sharp binary treatment*). Panels 4a and 4b plot the event-time coefficients from a version of equation 1.4, with the market-level number of unique users and total revenue as dependent variables, respectively. Following implementation, we can see a large and steady increase in the number of weekly users as well as in revenue for affected markets compared to unaffected markets.

These figures are also useful to examine trends in the development of market outcomes prior to integration. I find similar trends between eventual treated and control markets in the period prior to integration. These pre-trends serve as suggestive evidence that the key identifying assumption of the difference-in-difference estimator, parallel trends in the absence of treatment, holds in this settings.

Table 2 presents the formal estimation results of equation 1.4. Column (1) uses the binary treatment definition, above- and below-median treatment intensity. Column (2) uses the sharp binary treatment definition, as describes in Section 1.4.3; specifically, this result ignores markets that were only 'weakly' treated and considers only markets that were not affected by the integration as compared with markets where treatment intensity exceeded the 75th percentile. Column (3) uses the continuous underlying treatment intensity. Across the board, there are statistically significant (at 1 percent level) positive effects of increasing entry on all three measures of demand and total market size. We can see in Column (1) that a positive increase in the supply-side of the market leads to an increase of 36.4% in the number of users on the platform. To understand the magnitude of this effect, note that the mean percentage change in the share of businesses on YTP in treated markets is about 85% of the initial share. This outcome implies a demand elasticity of approximately 0.42 with respect to supply. The effect on the weekly number of orders is similar in magnitude (36.7%) and the effect on total revenue is slightly higher, 58.7%. As expected, Column (2) estimates, obtained by considering markets that received sharper treatment, are larger for all three outcomes. The estimate in Column (3) captures the effect of changing the share of businesses on the platform from 0% to 100%. In contrast, though, the mean increase in the share of restaurants on the platform following the partnership is only 2% (Table 1). Thus, to get a better understanding of the magnitude of the estimates in Column (3), we need to multiply the two numbers, suggesting an increase of about 20% in the number of users on the platform.

Taken together, these results support the main prediction of two-sided markets literature (Katz and Shapiro, 1985, Rochet and Tirole, 2006): An increase in the supply side of the market draws more consumers into the market and leads to subsequent increases in demand, quantity purchased, and total revenue. This finding supports the main forces described in the model and motivates the subsequent analysis.

1.5.2 Effect on Incumbent Firms

Average Effects After establishing that the supply-side expansion indeed increased total market size, I now turn to explore the main research question: How did entry of new firms affect incumbent firms' performance? Graphical results appear in Figure A2. Table 3 displays the estimation results of equation 1.4. I restrict attention solely to *incumbent businesses*, i.e., businesses that operated on YTP prior to the Grubhub integration. Table 3's structure is similar to Table 2, with Columns (1) and (2) displaying the binary and sharp binary treatment definition and Column (3) presenting the continuous measure.

Panel A showcases the effect of entry on the weekly number of orders per business. I find a null average effect of entry on the number of orders per business. The effect in Column (3) is statistically significant but is economically small, a little less than 1% in the average city.[21] In contrast, in Panel B, I find significant *positive* effects of about 4.5% on incumbent firms' revenue. The positive effects on weekly revenue are statistically significant and are robust to the specific definition of treatment.

Taken together, Table 3 finds weak but positive effects of entry on firm performance. This result suggests that the countervailing forces of market size and market share described in Proposition 1 offset each other, with the market-size effect slightly dominating. This finding would have been extremely difficult to explain without taking into account the expansion in market size.

Heterogeneous Effects While the average effects are consistent with the countervailing forces affecting firms, it is difficult to assess the influence for each force separately. To address this challenge, I test the second prediction of Proposition 1, wherein entry will differentially affect high- and low-quality firms. As discussed in Section 1.4.1, I use the relative rating within city on the eve of integration, including for the newly added firms, as a measure of restaurant quality.

Figure 5 depicts the differences in weekly performance between treated and control markets by restaurant quality. Panel 5a plots the event-time coefficients from a version of equation 1.4 using the weekly number of orders for *highly-rated firms*. Similarly, Panel 5b plots the event-time coefficients for *low-rated firms*. The figure paints a clear picture: The average effect masks considerable heterogeneities in the impact of entrance on incumbent firms. In particular, entry has opposite effects on high- and low-quality businesses. For high-quality firms, we see a clear upward trend following integration, leading to substantial increases in the number of orders per week and vice versa for low-quality firms. Figure 6

[21]The estimated effect is 38.3% and the mean treatment intensity is 2%: $38.3\% * 2\% \approx 0.8\%$.

presents similar and sharper results for the effect of weekly revenue. Finally, it is worth noting that in all of the figures above, we see similar trends between the eventually treated and the control markets in the period prior to integration. This observation provides further reassurance that the parallel trends assumption, central to the difference-in-difference design, holds even when restricting attention to the subset of high- or low-rated firms.

Table 4 presents the estimation results of equation 1.5 on businesses' number of weekly orders and revenue as outcomes. Columns (1) and (2) present the results when using the binary definition, columns (3) and (4) use the sharp binary treatment definition, and columns (5) and (6) use the continuous measure of treatment intensity. The odd columns—(1), (3), and (5)—use the above-/below-median definition of quality, and the even columns—(2), (4), and (6)—drop the interquartile range in terms of ratings. The $Treat*Post$ variable captures the effect of entry on high-quality incumbents. As we can see on Panel A Column (1), entry increases the number of weekly orders by 3.6%. The positive effect of entry on high-quality incumbents is consistent across specifications and ranges from 3.6% to 6%, all significant at the one-percent level. As expected, sharper treatment and quality definitions yield stronger results: The estimated coefficient is largest in Column (4) (sharp treatment, sharp rating) and smallest in Column (1) (weak treatment, weak rating).[22] The variable $Treat*Post*Low$ captures the differential effect of entry on low-rated compared to highly-rated incumbents. To get the total effect of entry on low-rated firms, we should add the coefficients for high and low types. This result is presented in the third line, marked as $\beta_1 + \beta_2$ (referring to the notation in equation 1.5). Estimates for low-quality sellers are the reverse mirror image of high-quality businesses: I find statistically significant *negative* effects of entry on low-quality incumbents, with the decreased weekly number of orders ranging from 2.6% to 5.4%. Similar to the effect on high-quality firms, these results increase in absolute magnitude as the treatment definition becomes sharper.

Panel B presents similar effects of entry on weekly firm revenue. Again, I find that entry leads to increases in weekly revenue for high-quality firms and vice versa for low-quality firms. The point-estimates on revenue are substantially larger than those on the number of orders: I find increases as high as 15.8% in weekly revenue for high-quality firms and drops of up to 9.2% for low-quality firms. These estimates are much noisier than in Panel A, and though always negative, the estimates on low-quality firms are statistically insignificant in Column (1) and Column (5). Nevertheless, the effects on high-quality firms and the difference in effects between high- and low-quality firms are consistently significant at the one-percent level across all specifications.

To conclude, the estimates in Table 4 complement the graphical analysis and support Proposition 1. The weak average effect masks considerable heterogeneity in firm performance following the entry of new competitors to the platform. The positive effects of entry are generated exclusively by the high-quality businesses. In contrast, low-quality firms are negatively affected by firm entry into the platform. While the highly-rated firms seem to benefit from the increased market size and the low-rated firms suffer from a decrease in market share, it would be incorrect to equate the magnitudes of these two forces with the

[22]Since the estimates using continuous intensity are hard to interpret and are generally similar to the results when using the binary treatment definition, in the following tables, I will focus attention on the binary- and sharp binary–treatment definitions.

reduced form estimates by quality level. The reason is that both high and low types are affected by both forces, though to different degrees. The heterogeneous effects for each firm type capture different mixtures of the market-size and -share effects. In order to estimate each effect in complete isolation, I must impose additional structure on the model, which I will discuss in more detail in Section 1.6.

1.5.3 Robustness of the Main Results

In this section, I briefly describe the robustness checks for the main results. A more complete description, as well as additional robustness tests, can be found in Appendix A.4.2. While the exact magnitudes of the treatment effects oscillate across different specifications, the main results are consistently robust to all of the alternative specifications discussed below.

Market Definition: Geographic area The main specification treats the city as the relevant market. However, this definition might be viewed as either too narrow (in large cities) or too broad (in clusters of small cities). To address these concerns, I reconstruct the data twice: first by narrowing the geographical definition of markets using the 5-digit zip code as the relevant market, and second by broadening market definition using the county as the relevant market. I find that the main effects are robust to the specific geographical definition of the relevant market. The results are presented in Table A3.

Market Definition: Food category The main specification considered only geographic boundaries in the market definition. One concern may be that other dimensions, such as food category, should be taken into account when defining relevant competitors. For instance, pizza restaurants in San Francisco might be competing with other pizza places in the city but not with sushi or Mexican restaurants. To address this concern, I reconstruct the data defining markets by city-category combinations. Potential pitfalls of this analysis are discussed in Appendix A.4.2. Nevertheless, the main results, presented in Table A4, are robust to the alternative market definition.

Treatment Definition So far, treatment intensity was defined as the change in the share of businesses on the platform. This definition, however, does not capture the change relative to the initial share in the market. I estimate equation 1.5 using:

$$TI = \frac{\text{\# restaurants on YTP Post - \# restaurants on YTP Pre}}{\text{\# restaurants on YTP Pre}}$$

which is the percentage change in the share of restaurants on the platform as the relevant treatment intensity. I find that the main results, presented in Table A4, are robust to the alternative definition of treatment.

Outliers There are substantial differences between markets in the sample: Some are small towns with only a few businesses and others are huge metropolitan areas. One potential concern is that high-leverage outliers drive the results. To address this concern, I perform two sets of tests: First, I estimate the main specification while excluding the 5% largest and smallest cities. Second, I generate p-values using randomization inference tests, which provide more robust and accurate inferences in the presence of high-leverage observations and when interacting the treatment assignment with unit characteristics (Young, 2016).[23] The main results, presented in Table A5, are robust to both of these tests.

[23]Robustness inference provides a test for exact hypotheses using the random assignment of treatment

Unbalanced Observables As mentioned in Section 1.4.2, there are significant differences in city characteristics across treatment and control cities. Treated cities are, on average, larger, have more restaurants, and have a higher share of restaurants on YTP. While this finding does not violate the identifying assumptions, I conduct several tests to verify that initial differences in market characteristics are not driving the results. First, I perform inverse probability weighting to account for the different probabilities of selection into treatment based on observables (Hirano et al., 2003) Second, I conduct a more demanding test, which takes advantage of the fact that treatment intensity is a continuous variable. In particular, I estimate the propensity score to receive treatment and then assign a binary treatment indicator within each propensity score bin, effectively changing the threshold for assignments into treatment as a function of the propensity score (a more detailed discussion appears in Appendix A.4.2). The main estimates, presented in Table A6, are robust to both of these tests.

1.5.4 Firm Response

In this section, I examine how firms respond to changes in market structure. I explore responses on three dimensions: pricing, investment in product quality, and advertisement.

Price Response I begin by studying how restaurants readjust prices in response to market expansion. Since restaurants sell multiple items, constructing relevant prices is not a trivial task. First, I can only estimate price changes for frequently ordered items. Second, I observe price paid rather than menu prices, which embeds noise in the price-response data, as dish modifications (e.g., "add chicken" or "make large") are not always documented. To address these issues, first, I restrict attention to the six most ordered items in each restaurant, and second, I develop an algorithm to separate true price changes from dish modifications. The specific details are discussed in Appendix A.3.

Table 5 presents the results. Columns (1) and (2) present the estimates of equation 1.4, with price as the dependent variable and item fixed effects. None of the coefficients on the average effects are statistically significant at the five-percent level. Moreover, the point estimates are generally economically small and have inconsistent signs across specifications. For instance, Column (1) finds that entry leads to an average increase of 4.4% in prices (this effect is significant at the ten-percent level), while Column (2) suggests a statistically insignificant decrease of 1.4% in average prices. Columns (3)–(6) present estimates of equation 1.5 using the different definitions of treatment assignment and high-quality. While the effect on prices in highly-rated restaurants seem to be generally positive and the effect on prices in low-rated restaurants are generally more negative, none of the coefficients on the heterogeneous effects are statistically significant and all are economically small.

To conclude, I do not find evidence that the entry of new businesses affected the prices

instead of the (asymptotic) distribution of the error term. Intuitively, the estimation procedure iteratively reassigns units into treatment and control and estimates the treatment effect at each iteration. Then, p-values are calculated from the location of the true estimates in the distribution of estimates from potential treatment allocation. Since randomization inference is robust to small sample sizes, this approach is mostly recommended for analyzing experiments (Athey and Imbens, 2017). Nevertheless, Young (2016) shows that this methodology also performs better in settings with high-leverage observations or when interacting the treatment with unit characteristics.

of incumbent restaurants. The null effect may be an artifact of the specific setting and data limitations: First, restaurants' prices are notoriously sticky and are often given as an example of businesses with high "menu cost" (Hobijn et al., 2006, Bils and Klenow, 2004, Zbaracki et al., 2004). Empirically, I find little price variation over time, as less than 20% of *all* food items in the sample have one or more price changes in a period of two years. Second, interviews conducted with YTP employees and delivery services suggest that online and offline prices rarely differ. Anecdotal evidence[24] as well as the relatively low weekly revenue from orders suggest that delivery and takeout constitute a relatively small portion of firms' profit function and thus are not given much weight in setting prices. Finally, the data limitations discussed above may be restricting this study's ability to detect price changes even if they are present in the data—when attempting to identify true price changes, I make quite a few assumptions and drop a significant portion of available data.

Investment in Quality The main analysis shows that entry increases revenue for high-quality sellers and decreases revenue for low-quality sellers. These findings directly imply that entry increases the *return to higher ratings*. Figure 7 demonstrates the intuition graphically. Panel 7a presents the treatment effect by rating decile. The change in revenue is monotonically increasing in rating decile and is negative for low-rated restaurants versus positive for high-rated ones. Panel 7b shows the relationship between the weekly number of orders and rating decile before and after the integration. As we can see, there is a strong positive relationship between sales and ratings prior to integration. Following integration, the trend line becomes even steeper, suggesting that moving up rankings increases sales by more than what would happen in the term prior to integration. As the return to high quality grows, so do the incentives to invest in quality. Accordingly, if restaurants can (at least partially) affect their ratings, then we would expect to see subsequent increases in firms' ratings in treated markets.

To test this intuition, I estimate the effect of entry on the flow of incumbents' weekly ratings. Table 6 presents the estimation results of equations 1.4 and 1.5, with weekly ratings as the dependent variable. Columns (1) and (2) present the average effect and columns (3)–(6) present the heterogeneous effects by high- and low-rated restaurants. I find a small but statistically significant effect of integration on subsequent ratings, ranging between 0.6%–0.8%. The effect seems to be mostly driven by high-quality firms, though the differences between firm types are not statistically significant and the point estimates are sometimes positive. The positive effect on high-rated firms is consistent across specifications and exceeds 1% for the sharpest treatment and rating definitions.

To get a better understanding of the size of these effects, note that an increase of 1% in ratings will move the median business up by about 3 percentiles in the ratings distribution.[25] Using the results from Table 4, a back of the envelope calculation suggests that, for the median restaurant, a 1% increase in ratings percentile is associated with a little less than 1% increase in weekly revenue. The relatively small magnitude is unsurprising: Ratings on Yelp are a combination of reviews for delivery services as well as for the brick and mortar restaurant. It is unclear whether restaurants have much room to improve ratings, and even if

[24]For instance, a Morgan Stanley report finds that online food delivery comprises only 6% of the total restaurant market.

[25]In equilibrium, however, all firms will invest in quality, resulting in a red queen's race: all firms increase investment just to stay at the same rating percentile.

so, whether this effort will have a meaningful impact on their ratings. Furthermore, reviews are relatively rare and thus extremely noisy; for the sample of incumbent restaurants, the median number of reviews per week is zero and even the 75th percentile is only one weekly review.

A potential concern is whether the estimated effects reflect true changes in quality or merely changes in rating behavior. First, results may be driven by rating inflation (Horton and Golden, 2018, Nosko and Tadelis, 2018). This concern is mitigated by the test conducted to support the identifying assumption: I test for changes in ratings trends for *non-YTP* businesses, and find null (and slightly negative) effects of entry into YTP on subsequent ratings of non-YTP businesses (Table A2). Nevertheless, to address concerns of differential rating inflation across markets, Appendix A.4.3 and Table A11 present a placebo specification in which integration is counterfactually coded at the middle of the pre-treatment period. The placebo test yields null results, suggesting that the effects are not driven by differential trends in rating inflation.

The second concern is that selection into specific services is correlated with rating behavior (Kovács and Sharkey, 2014, Fradkin et al., 2018). For instance, if users who use delivery services also tend to rate more leniently, then increases in online ordering might mechanically drive up the ratings of restaurants. To alleviate this concern, I tests for differential changes in raters' leniency, which I define as the average rating across all reviews posted by the user (Table A11). I find no significant differences between raters' leniency in treated or control markets following the integration.

Advertising The previous section examined how changes in market size affect firms' incentives to invest in quality. I now study an alternative way for firms to attract consumers. Specifically, Yelp offers multiple services to improve firms' appeal, with the two most common services being profile enhancement, which is a bundle designed to increase the attractiveness and conversion rate of the Yelp business page, and targeted ads. A more detailed description of the services Yelp offers to businesses is presented in Appendix A.3. In my analysis here, I use data on the total weekly revenue collected from businesses by Yelp, which I refer to simply as advertising expenditure.[26] Since treatment is administered by city, I restrict attention to campaigns purchased at the local-level, excluding national- or franchise-level campaigns.

Table 7 presents the results. Columns (1) and (2) present the average effect of market expansion on firms' advertisement spending. I find that increased entry leads to a 2.7%–3.5% *decrease* in total firms' spending on advertisements. Columns (3)–(6) decompose the effect by firm quality and find that the drop is concentrated among high-quality firms, which spend 3.4%–8.1% less on advertising following integration. In contrast, I do not find any consistent or statistically significant changes in spending among low-quality firms. The same patterns hold when examining the fixed and variable revenue separately, though the results on variable revenue are stronger both in terms of magnitude and statistical significance. Note that, in the period before integration, high-quality firms spent as much as 45% more on advertising than low-quality firms (consistent with Armstrong et al., 2009). Thus, though entry decreases the advertising gaps, high-quality firms still advertise more than low-quality firms in the same market.[27]

[26]Decomposition by revenue source can be found in Table A12.

[27]This finding does *not* have a casual interpretation, since rating and advertising behavior are not randomly assigned. For example, it is unclear whether high-rating leads to higher spending or whether more

This finding suggests that firm advertising and platform size expansion acts as substitutes; The increased sales generated by the growth in market size crowd-out the investment in advertising.[28] One possible mechanism is increasing marginal costs of production or, in the extreme case, capacity constraints. Since I do not find any significant effect on prices, if per unit costs are increasing, then the net profit per unit is decreasing in the number of units, leading to a decrease in the marginal benefit of selling an additional unit. This consideration implies a decrease in the returns to advertising for high-quality firms. Intuitively, if firms have capacity constraints (or infinite marginal cost) then, once they reach capacity, there are zero returns to additional advertising.

1.5.5 Mechanisms

The Importance of Differentiation While the main analysis focuses on vertical differentiation (quality), horizontal differentiation, or the similarity between incumbents and entrants, may be an important determinant of incumbents' outcomes. Specifically, I study how the overlap in the food category of entrants and incumbents effects incumbents' performance.[29] The direction of the effect is theoretically ambiguous and relates to the forces described above: On the one hand, close alternatives compete more fiercely with incumbents, strengthening the negative *market share* effect (Hotelling, 1929, Barney, 1991). On the other hand, in a world with heterogeneous consumers, close alternatives can attract consumers who are more interested in the specific food category, increasing the positive *market size* effect.

I examine whether incumbents perform better or worse when a larger share of added restaurants are in the same food category. In particular, I regress a version of equation 1.4 in which I restrict attention in the analysis to cities that were affected by the partnership, and define treatment as the share of restaurants of the same food category out of the total number of entrants. For instance, if 100 new businesses joined YTP and 30 of the entrants were pizza places, then an incumbent pizza place will receive a treatment of 30%.

The results are presented in Table 8. In the odd columns, treatment is an indicator for whether any businesses of the same food category joined YTP, whereas the even columns use the continuous share of restaurants of the same food category out of the total number of entrants. The estimated effects of similarity on incumbents' performance are negative in all specifications and are as low as –2.9% and –6.4% on the percentage change in the number of orders and revenue, respectively. The estimates, however, are only marginally statistically significant. Taken together, these results suggest that differentiation helps maintain incumbents' competitive advantage and mitigates the deleterious effects of entry.

Is YTP Cannibalizing Revenue from Other Platforms? Table 4 shows that high-quality restaurants experienced increases in both weekly sales and revenue following the addition of new restaurants to the platform. While it is clear that integration positively affected the revenue of high-quality restaurants on YTP, the effect of integration on the total revenue of high-quality restaurants remains unclear. In particular, the new consumers

advertising leads to better reviews.

[28]This result is consistent with Hollenbeck et al. (2019) who firnd that demand generated by higher ratings substitutes advertising expenditure.

[29]I focus on the 21 largest food categories, which consists of about 87% of all observations. See discussion in “Market Definition: food category” in Appendix A.4.2 for details.

making orders on YTP might be substituting away from other forms of interaction with the restaurant, such as other delivery services.

It is hard to fully resolve these concerns since I do not have data on total firm revenue nor do I see consumers' transactions on other platforms. Nevertheless, note that in is difficult to ascribe the differential impacts by firm quality to cannibalization of other channels. In particular, such explanation would only work if platform expansion causes more cannibalization for high-quality firms compared to low-quality firms, which seems unlikely. In addition, I take advantage of the detailed search data on Yelp to generate evidence suggesting that the increase in orders from YTP is not cannibalizing other sources of revenue. Specifically, I examine whether users substitute away from ordering on other platforms. If incumbents offered delivery through other platforms, it may well be the case that, as YTP becomes more attractive, consumers leave the old platforms and switch to YTP. In this case, the total number of orders remains the same for a given restaurant, but the number of orders on YTP increases. This concern can be stated as follows: Consumers already know which restaurant to order from, and are merely selecting the channel to do so. I argue that if this situation is the case, we can expect to see changes in search patterns: We would expect consumers to search less and enter the order menu more quickly.

Table 9 presents the results of the estimating equation 1.4 at the city level, with search metrics as the outcomes. The sample includes only sessions in which a user ordered from a restaurant for the first time. Details regarding the construction of the sample, as well as limitations of this approach, are described in Appendix A.3. Panel A presents the effect of treatment on the average number of searches (search queries entered) prior to ordering. I find a weak *positive* effect on the number of searches, suggesting that integration caused users to search more intensively. Panel B presents the effect of treatment on the number of business pages viewed by users. The effects are economically small and insignificant. Panel C presents the total time spent on the platform. I find a significant and substantial *increase* in the time spent on the platform, ranging from 35% to 50%. Panel D presents the effect on the Levenshtein distance between the first search query and the name of the restaurant eventually chosen. The Levenshtein distance is defined as the number of character changes needed to move from the query to restaurant name, i.e., lower numbers imply a narrower search; if consumers already know which restaurant they want to order from, then we can expect the distance between search queries and the selected restaurants to decrease following the partnership. I find that average Levenshtein distance decreases by 3.3% to 5%, suggesting that users do use more specific queries. This result, however, is misleading; even after the reduction, the median distance in the sample is 12, which is so large that it is unlikely that consumers have a good idea of what they are looking for when beginning the search.[30] Moreover, almost 80% of the queries include a generic search term such as "Delivery," "Chinese Takeout," or "Thai food." To examine whether integration affects the use of these generic search queries, I estimate a linear probability model with an indicator for a generic search as the outcome. The results are presented in Panel E; I find no evidence of reduction in generic searches in treated markets following integration.

[30]To get a sense of the magnitudes, the Levenshtein distance between the search query "Pizza" and the fictitious restaurant "Oren's Pizza" is only 7. More alarming, the distance between "Chinese food" and the same restaurant is only 10!

Taken together, these results suggest that search intensity is not decreasing following integration and even seems to be slightly increasing. This observation is consistent with users ordering from restaurants for the first time, as opposed to simply changing the delivery platform, suggesting an that the increase in revenue on YTP represents an increase in total revenue as well.

Are Consumers Responding to Firm Quality or to the Ordering of Search Results? The model argues that the main effects are driven by consumers' selection of higher quality. A potential alternative explanation for the results is that the order of search results is affecting consumers' choices. In particular, if Yelp ratings are highly correlated with search results' sequence, then entry may be mechanically decreasing sales of low-quality business by reducing their salience in the search results and vice versa for high-quality types.

I first explore the importance of Yelp ratings in determining the order of search results.[31] I find that while high-rated business are usually ranked slightly lower (we expect the relationship to be negative since lower ranks appear first), the correlation between Yelp rating and search result rank is smaller than -0.1. Moreover, on average, improving the star-rating from two to five stars improves rankings by less than three ranks. In comparison, the mean rank in the YTP data is 18, and the difference between the average low- and high-quality businesses is about one star (3.5 and 4.5). Taken together, these results imply only a weak relationship between ratings and search results' orders.

Second, I formally test whether controlling for search results' orders changes the main results. To this end, I construct an index of the average weekly search-result ranks of a business, including all searches in which the business appears. As expected, the mean ranking for low-quality businesses is slightly lower than high-quality businesses, 15.5 and 18.5, respectively. I then re-estimate the main specification, flexibly controlling for the average weekly ranking.[32] The results are presented in Table 10. I find that the main results are robust to the inclusion of average search results' ranks. All coefficients are significant at the one-percent level and are slightly more negative compared to the main specification, which implies a stronger effect on low-quality types and (slightly) weaker effects for high-quality types. I thus conclude that firms' quality is important even when controlling for the ordering of search results.

1.6 Structural Estimation

The reduced form results imply that entry has a positive impact on the sales and revenue of high-quality firms and a negative impact on low-quality firms. The main constraint of the reduced form analysis, however, is that these results are only relevant to a small segment of potential firm entry onto the platform. In particular, the median city in the sample has less than 5% of restaurants in the city on YTP even after integration. It is unclear if the results carry over to situations in which a larger percentage of the population of restaurants already participates in the market. For example, we can expect the market-size effect to

[31] Yelp's search results ordering is proprietary, and I was not able to learn about the specific characteristics determining the orders or the relative weights given to each attribute. The observations I make are based solely on analysis of the relationship between firms' characteristics and their relative ranks in search results.

[32] Formally, I estimate equation 1.5 with a third-order polynomial of the weekly rank index.

have decreasing returns, i.e., the magnitude of additional increases in market size decreases when a large number of businesses are already on the platform. To study how entry affects firm performance as market participation grows, I impose additional structure on the model and the data-generating process. The structural model allows me to perform out-of-sample predictions for the full schedule of potential entrants.

1.6.1 Preliminaries

Setup I begin with the second stage of the model: Conditional on using the platform, consumers' decisions follow the standard discrete choice model (McFadden et al., 1973, Berry et al., 1995). Specifically, I assume that the indirect latent utility of consumer i from buying product j in market t is:

$$U_{ijt} = \beta_i X_{jt} - \alpha_i P_j + \delta_t + \epsilon_{ijt} \tag{1.6}$$

Where X_{jt} are observed product characteristics, and P_j is the restaurant's price range as captured by the Yelp Dollar Rating. δ_t are the combined statistical areas' (CSAs') fixed effects, and ϵ_{ijt} represents the random horizontal utility shock, assumed to be distributed i.i.d. extreme value type 1. The random coefficients, β_i, are consumer-taste parameters for different product characteristics, and α_i captures the disutility from price. I allow α_i to vary across consumers according to:

$$\alpha_i = \alpha + \Pi D_i + \sigma_v v_i, \quad v_i \sim N(0,1) \tag{1.7}$$

Where D_i represents observed consumer characteristics, Π is a matrix of coefficients that measures how taste characteristics vary with demographics, v_i are unobserved shock to preferences assumed to be normally distributed, and σ_v is the variance. Consumer characteristics, D_i, include gender, age, and income level. Restaurant characteristics include restaurant's food category, Yelp rating, and mean income at restaurant location. I also include the quintile of a restaurant's rating in the ratings distribution for the market (*ranking*) and the ranking of restaurant's rating out of the restaurants in the same food category for the market (*category ranking*).[33]

Moving to the first stage of the model, I assume that consumers' entry costs are normally distributed and allow the distribution to vary by (potential) consumers' income.[34]

$$H(c) \sim N(\mu_i, \sigma), \text{ with } \mu_i = \mu + \gamma D_i \tag{1.8}$$

[33]A few additional notes on the functional form of the utility function: 1) The utility function does not include business fixed effects since there are over 30,000 unique businesses in the final sample and only a handful of observations per business. In addition, I observe little-to-no variation over time in the covariates of interest. I use fixed effects by CSA, since this is the set from which I draw competitors in the simulation. 2) The formulation in 1.6 is more flexible than described in Section 1.2. The latter only allows for vertical and horizontal differentiation in consumers' taste. The random coefficients specification also allows for a mixture of the two. 3) To address the concern that firms readjust prices in response to unobserved demand shocks, I instrument for dollar rating using BLP instruments (Berry et al., 1995). These instruments are likely to work well in this setting: First, as discussed in Section 1.5.4, I rarely see any changes in item prices, even after large shocks to the market. Second, entry of new firms is plausibly exogenous in this setting, and thus entering firms' characteristics satisfy the instrumental variable exclusion restriction.

[34]First, there is no reason ex ante to assume that entry cost will follow the normal distribution, especially given the fact that we expect cost to be strictly positive. Nevertheless, I experimented with several alternative

Empirically, potential market size is defined as the number of households in the city. The number of users interested in the platform is defined as the number of unique users in a given market searching for variations of the words "delivery," "takeout," or "pickup," or using any of the YTP filters.[35] The share of consumers on the platform is given by the quintet of the above. Finally, I normalize the utility from the outside option, ω, to zero. A detailed description of the data used and the construction of variables for the structural model can be found in Appendix A.3.

Estimation Estimation begins at the second stage and follows the methods developed in Berry et al. (1995) and Nevo (2000). For this part, the market share of the outside good is measured as the share of consumers who do not order on the platform out of the number of consumers *searching on the platform.* I use the contraction mapping theorem and GMM estimation developed in Berry et al. (1995) to estimate the linear and non-linear parameters of the utility function, $\theta = \{\alpha, \beta, \Pi, \sigma\}$.

Given the distribution of ϵ, once we identified the parameters of the utility function, the expected utility from the set of restaurants available in the markets is given by: [36]

$$E[\max_j u_{ijt}] = \int_{D_i} log(\sum exp\{u_{ijt}\}) \tag{1.9}$$

Since ω is normalized to 0, consumers only use the platform when $E[\max_j u_{ijt}] > c_i$, which happens with probability $\Phi(E[\max_j u_{ijt}])$. To identify the parameters of the entry-cost distribution $\theta_2 = \{\mu, \gamma, \sigma\}$, I use the estimated utilities and the empirical share of users searching on the platform, S_{jt}, to derive the minimum distance estimator:

$$\theta_2 = \arg\min_x \left\{ (\Phi(E[\max_j u_{ijt}]; x) - S_{jt})' \ \hat{\Lambda} \ (\Phi(E[\max_j u_{ijt}]; x) - S_{jt}) \right\} \tag{1.10}$$

Where $\Phi(\cdot)$ is the cdf of the normal distribution and $\hat{\Lambda}$ is the (empirical) efficient weighting matrix.

The fundamentals of the models are then used to simulate markets in which the percentage of firms on the platform out of total number of restaurants in the market increases from 1% up to 100%. The main outcomes of interest are consumer welfare, total market size, and firm performance under different market conditions. Since ratings play a crucial role in determining firms' outcomes, the results are presented by rating quintile. The simulation algorithm and estimation details are described further in Appendix A.5.

1.6.2 Results

Model Parameters Panel A of Table 11 presents the estimates of the utility function parameters. Column (1) does not use an instrument for price ratings and presents the

distributions, including log-normal, gamma, and exponential distributions, and the normal distribution seems to best fit the data. Second, income-levels are indicated by three levels (see Appendix A.3 for details). Formally, I model the mean in CSA j as: $\mu_j = f_{j1} * \mu_1 + f_{j2} * \mu_2 + f_{j3} * \mu_3$, where f_{jn} and μ_n are the share of the population that has income in the n-th bin, and estimated the mean for the n-th group, respectively.

[35]For a discussion of the potential limitations of this approach, see Appendix A.3.

[36]Formal proof for the simple discrete choice model is presented in Small and Rosen (1981).

estimates from a simple logit model without user heterogeneities in the utility function. Column (2) presents the estimates from a logit model without user heterogeneities, but instruments for dollar rating. Column (3) introduces a random coefficient on the dollar rating, and Column (4) adds consumer demographics to the random coefficient on the dollar rating.

First, when instrumenting for dollar ratings, higher dollar ratings, or higher average restaurant prices, lead to lower utility. Note that while the mean effect in Column (4) is positive, the mean effect in the population (taking into account the impact of demographics on the price coefficient and integrating over their respective prevalence in the population) is approximately -4.5. Second, both relative ranking and absolute rating affect consumer choices. As expected, ratings have a significant positive effect on consumers' utility. Ranking within category appears to have a stronger negative effect than general ranking. Conversely, ranking quintile (mostly) has an unexpected positive sign—though this finding should be interpreted cautiously, since ranking, ranking within category, and rating have strong correlations. When removing absolute rating or ranking within category from the model, the effect of total ranking becomes negative (not reported). All specifications also includes CSA fixed effects and food category dummies, which are not reported for brevity. Finally, as presented in Column (4), the estimates of the dollar rating coefficient shifters are extremely noisy, and none are statistically significant. For this reason, the results from Column (3) serve as the main specification, and the analysis using the estimates from the full model is presented in appendices A.1 and A.5.

Panel B presents the estimated parameters of the entry cost distribution function, θ_2 in equation 1.10. In Columns (3) and (4), I allow the mean of the entry cost distribution to differ by income level and present the weighted average across demographics. The estimate mean, μ, and standard error, σ, are approximately 0.2 and 0.09, respectively.

Simulations Table 12 presents the simulation results from the random coefficient model without demographics.[37] In the baseline model only 5% of the firms in the city are available on the platform. Columns present the relative change in outcome when moving from 5% to 10%, 20%, and so forth. The first row presents the estimated change in welfare when the percentage of firms on the platform increases. Welfare is monotonically increasing in the percentage of firms on the platform, but at a decreasing rate. For instance, increasing the percent of firms available on the platform from 5% to 20% increases welfare by more than 80%. However, to get a similar increase starting at 20%, the percentage of firms on the platform has to grow to about 80%, almost four times the change, in order to have the same impact on welfare. The pattern of decreasing returns is presented graphically in Figure 8. The same calculation can be used to derive the actual welfare change from YTP's partnership with Grubhub. For the median treated city in the sample, the percentage of firms on the platform increased from 3.5% to 5.5%. These magnitudes imply an average welfare increase of 32.15%.

The second row presents the changes in market size as a function of the fraction of firms on the platform. Again, market size is monotonically increasing in the fraction of firms, but at a decreasing rate. This result is unsurprising, as the share of consumers depends on

[37] The equivalents of Table 12 and Figure 9 from the simulation with demographics are presented in Table A13 and Figure A4 and described in Appendix A.5.

a concave transformation (through the distribution of entry cost) on the change in welfare. For the median treated city in the sample, this simulation result implies an increase of 27.5% in market size. For comparison, the reduced form estimates presented in Table 2 find an increase of approximately 35% in the number of unique users and weekly orders. The similar magnitudes suggest that the model does a fairly good job in fitting the data, and is, in fact, underestimating the impact of market size growth.

The second part of Table 12 presents the main results, namely, the change in firms' performance by rating quintile. A more convenient way to understand the results is graphically, using Figure 9. The horizontal axis details the percentage of firms participating in the platform (1% to 100%), and the vertical axis present standardized sales. Each gray dot represents average sales for rating quintile over 1,000 simulations, the lines are the smoothing splines for each rating quintile, and the stars mark the maximum point of each smoothed trend-line. The table and figure reveal several results: First, though not imposed by the estimation algorithm, for relatively low participation rates, the simulation results are consistent with the reduced form analysis: For the median treated city in the sample, firms in the lowest rating quartile lose about 4.8% in sales and firms in the highest quartile gain 5%. In comparison, the simulation predicts a loss of 5.1% for firms in the lower-rating quintile and a gain of 5.2% for firms in the highest quintile. Second, for every percentage of firms on the platform, higher-rated firms perform strictly better. Graphically, the curves never cross: the highest-rated firms always sell more than the second highest, who sell more than the third highest, and so forth.

Third, except for firms in the lowest-rating quintile, for which sales are strictly decreasing in percentage of firms on the platform, there is a non-monotonic relationship between the percentage of firms on the platform and sales. In particular, initially, sales grow with the percentage of firms on the platform. As the percentage of firms continues to grow, however, the trend changes and the effect of market competition starts to dominate. Thus, sales start declining for all types of firms when the percentage of firms on the platform exceeds 50%. At 100% participation, for example, sales are lower for all firms as compared to 5%, except for the highest-rated ones. This finding is consistent with the concavity of the market-size effect: as the percentage of firms grows, additional firms are less successful in attracting more consumers into the market but continue to compete with the incumbents. Finally, the bliss points, the points at which sales are maximized, are monotonically increasing in relation to firms' rating quintile. The bliss point is zero for lowest-rated firms, and increases to 3%, 11%, 14%, and 46% for firms in the 2nd, 3rd, 4th, and 5th quintile, i.e., the highest-rated firms generate the highest sales when 46% of the firms participate in the platform.

1.7 Conclusion

To conclude, this paper studies an important yet unanswered question regarding the impact of new-firm entry on incumbent firms in two-sided markets. Collaborating with Yelp Transactions Platform, I investigate how entrants bring new value to the platform and expand total market size while—in parallel—threatening the performance and market share of incumbent firms. Using a difference-in-differences research design, I study the relative magnitude of these forces empirically and find strong evidence of network externalities, i.e., that supply-side growth led to increased demand and total platform usage. In addition, the entry of new firms benefited incumbent firms, but the positive effects are concentrated solely around high-quality firms; low-quality firms, in contrast, experienced a reduction in both sales and revenue following entry. I also find evidence suggesting that firms respond to the changes in the competitive environment by increasing their investments in quality and by adjusting their advertising behavior. Finally, using a simple structural model, I extrapolate the results to additional market settings and find that, in general, higher-quality firms prefer larger markets with more firms.

The main contribution of the paper is to highlight the importance of firm interactions and competition *within* the platform. The results described in this study have important implications to firms operating in platform settings, to platform strategy, and to competition policy. First, I find that the effects of entry on incumbent firms are positive on average, but critically depend on firm characteristics (i.e., quality) as well as platform maturity and size. Incorporating these considerations into the analysis yields important insights regarding how firms operate in a platform setting: For example, the main findings of this paper suggest that the platform environment and firm identity will mediate the benefits to incumbents from setting barriers to entry on platforms, with different platform characteristics and firm qualities dictating the direction and magnitude of such strategies . These new considerations should be taken into account when determining incumbents' responses towards the threat of entry in platform markets. Consequently, future research would benefit from unpacking the question of whether firms internalize the positive spillovers generated by potential entrants and whether they differentially invest in deterring or even promoting the entry of new firms.

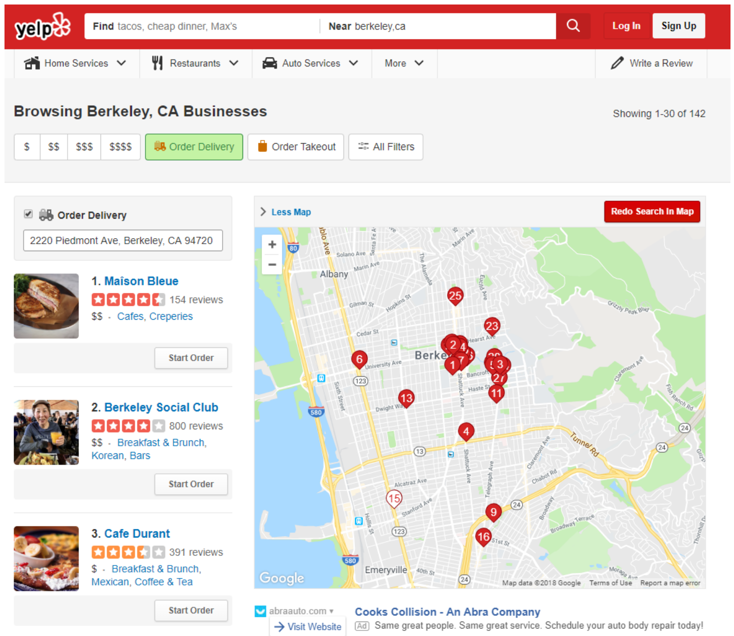
One limitation of the described analysis is that I only observe performance on one platform. While I present suggestive evidence that increased revenue on the platform is not strictly cannibalizing other sources of revenue, the model and empirical analysis generally abstain from addressing cross-platform competition and substitution patterns. A natural extension is to expand the analysis into additional datasets which include both multiple platforms and the larger market setting.

Second, expanding the number of suppliers on the platform naturally benefits the platform and is usually a central concern to platform managers. This paper contributes to the existing body of knowledge by demonstrating an additional channel by which expansion of the supply side may benefit the platform. Entry increases variety, creates additional value to consumers, and increases total transactions volume. At the same time, new firm entry raises the average quality on the platform: First, the volume high-quality sales grows while that of low-quality sales diminishes. Together, those imply that the average quality purchased by consumers is also increasing. Second, as the return to high quality increases, firms increase subsequent investments in quality. Finally, the simulations performed within this study further suggest that larger platforms are more attractive to high-quality firms, which subsequently signifies that the quality of the average entrant will improve as the platform grows. The current model studies only the short-term equilibrium and does not assess entry decisions, since entry into the market is quasi-random in our empirical setting. Nevertheless, these three mechanisms together suggest that entry does not just increase the size of the platform but also makes the average quality of the platform's firms better.

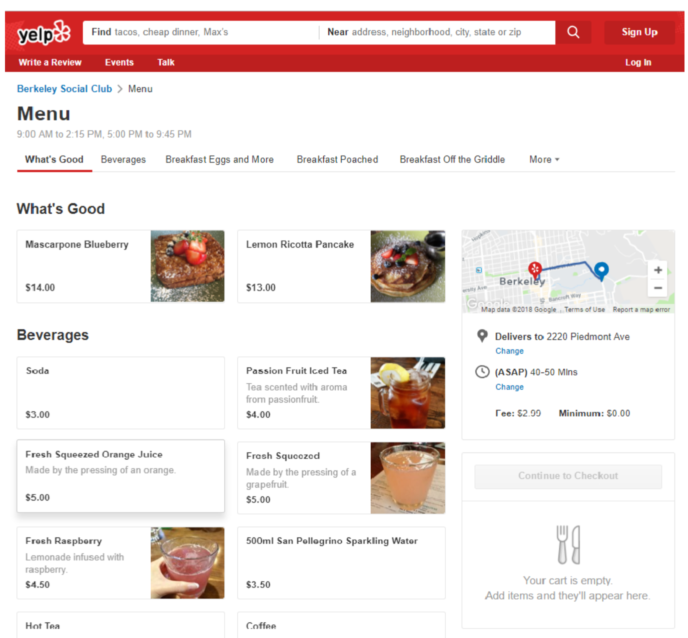
Lastly, policy-makers and regulators frequently direct attention to competition between platforms and the importance of restricting firms. This paper contributes to our understanding of the benefit and pitfalls of excessive regulation of platform markets. While addressing general impact on the market via the existence of multiple platforms is beyond the scope of this paper, the main results raise several important considerations that need to be incorporated into the decision-making process. In particular, restricting platform growth curtails the benefits detailed above—such as loss of positive network effects—and reduces competitiveness and quality on the platform. These negative implications must to be carefully examined as to not harm the businesses operating on the platform, the platform itself, and even the consumers that the regulators seek to protect.

Figures

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



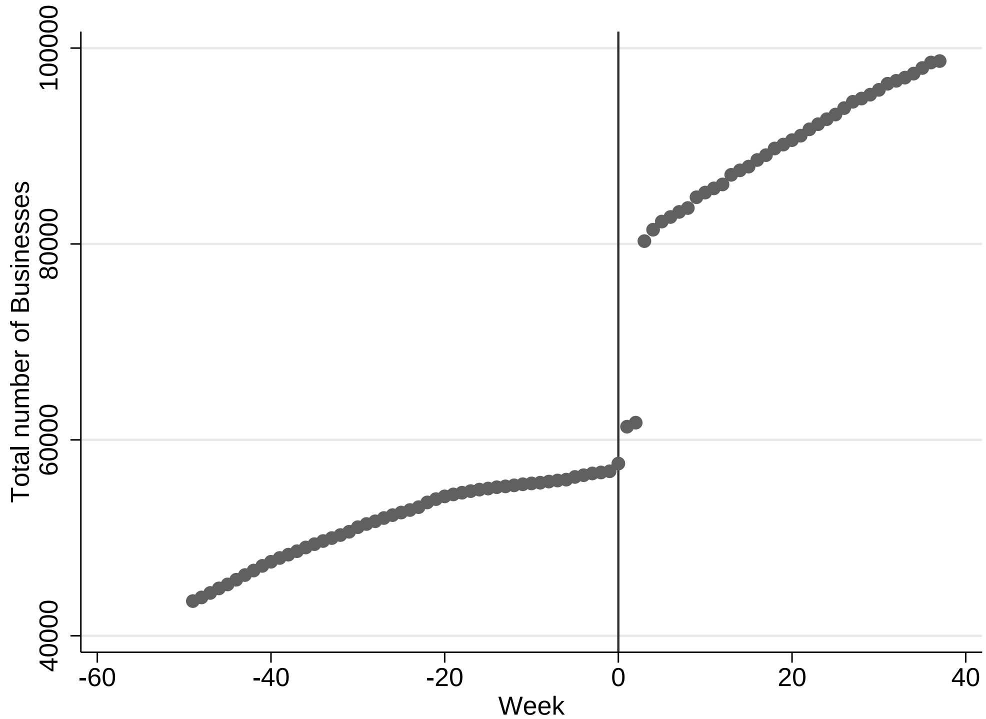
(a) Search on Yelp Transactions Platform



(b) Ordering on Yelp Transactions Platform

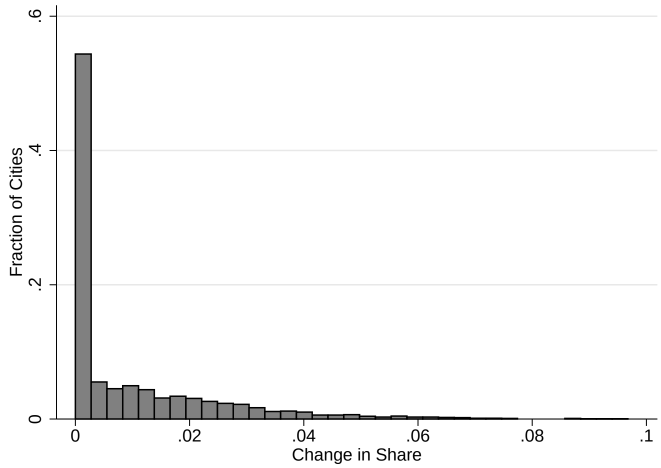
Note: Panel A presents the search results on YTP around Haas School of Business. Panel B presents a menu for a restaurant affiliated with Grubhub (as indicated in the bottom-right corner).

Figure 2: The Number of Businesses on YTP over time

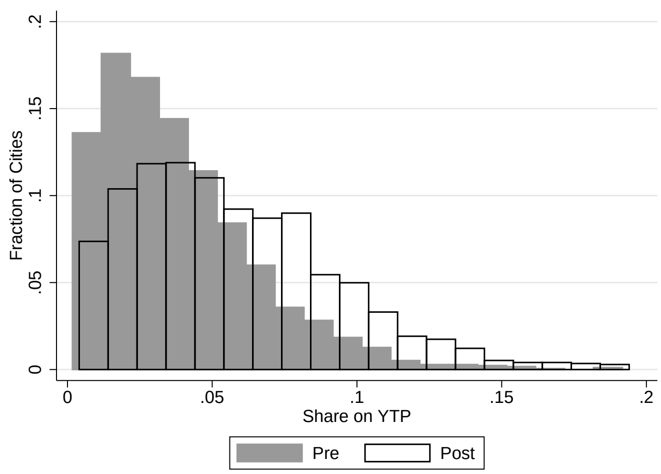


Note: The figure presents the development of the total number of restaurants available on YTP over time. The week of implementation is normalized to zero.

Figure 3: The Distribution of Treatment Intensity by City



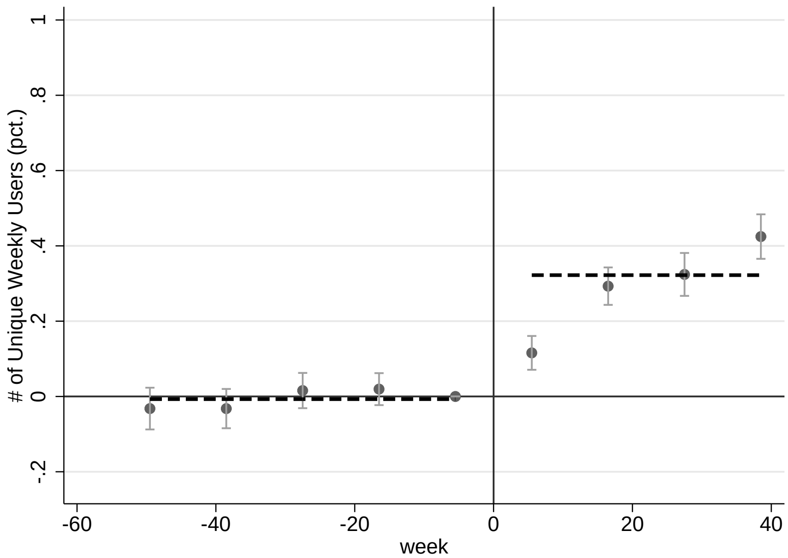
(a) Change in the Share of Businesses on YTP by City



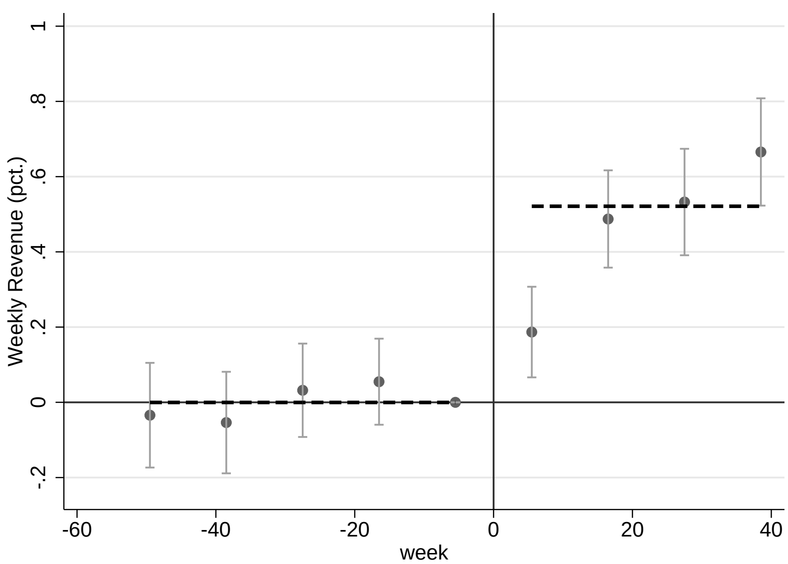
(b) Share of Businesses on YTP by City
(Conditional on Change)

Note: The figure presents the change in the percentage of restaurants available on YTP by city. Panel A presents the distribution of the percentage change by city. Panel B presents the distributions of the percentage of restaurants available on YTP before and after the partnership, for cities that were affected by the partnership.

Figure 4: Impact of Entry on Market-Level Outcomes (Percentages)



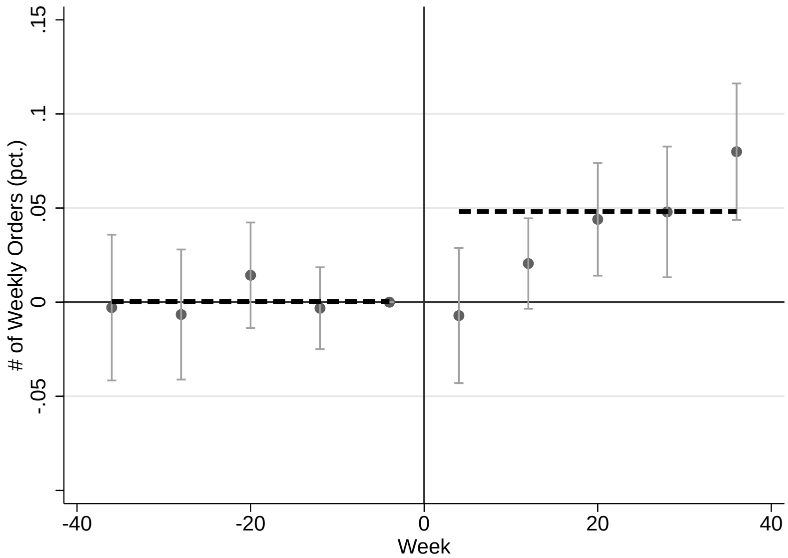
(a) Weekly Number of Unique Users



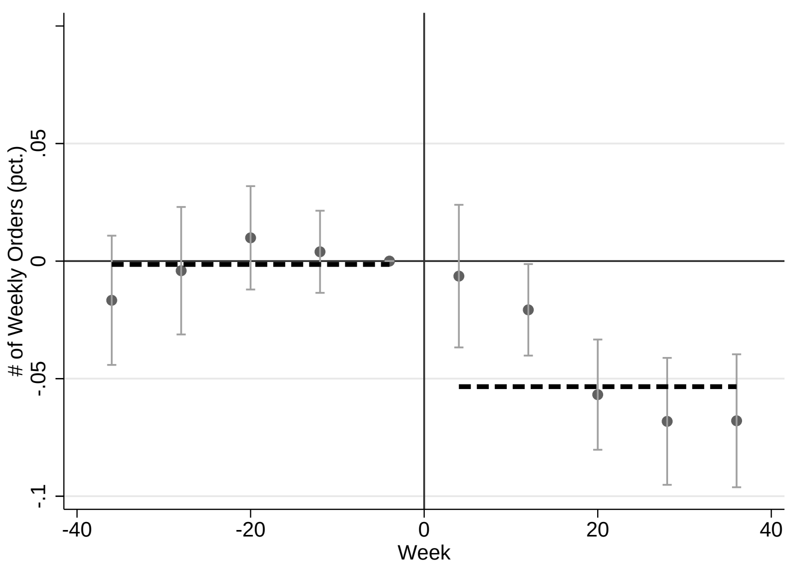
(b) Weekly Revenue

Note: This figure presents event-time estimates from a version of equation 1.4. The dependent variables are the inverse hyperbolic sine transformation and should be interpreted as percentage changes. The unit of observation is city-week, including both incumbent and newly added businesses. The dots represent point estimates from regressing the dependent variable on a treatment indicator interacted with nine-week bins, and city and week-state fixed effects. The treatment indicator compares cities that experienced almost no change in the percentage of businesses available on the platform to cities that experienced meaningful changes. The coefficient in the first period prior to implementation is normalized to zero. The vertical bar represent 95% confidence intervals, where standard errors are clustered at the city level.

Figure 5: Impact of Entry on Weekly Number of Orders Per Business by Firm Quality (Percentages)



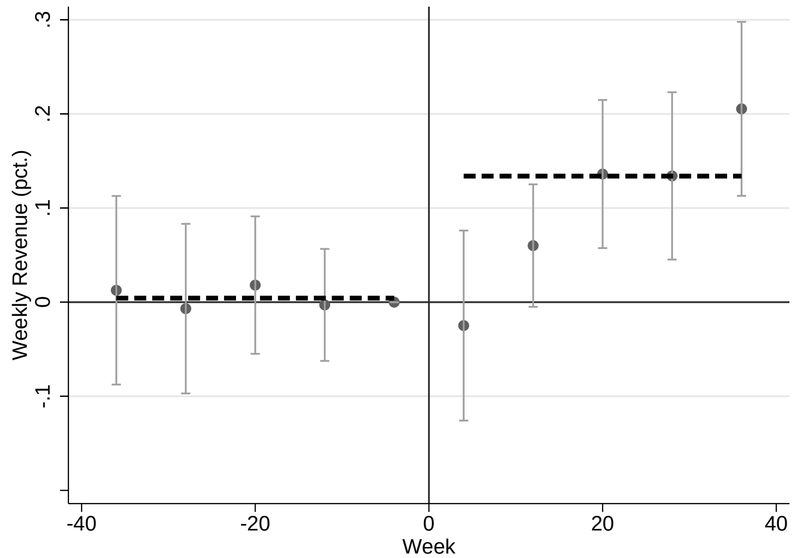
(a) High-rated Firms



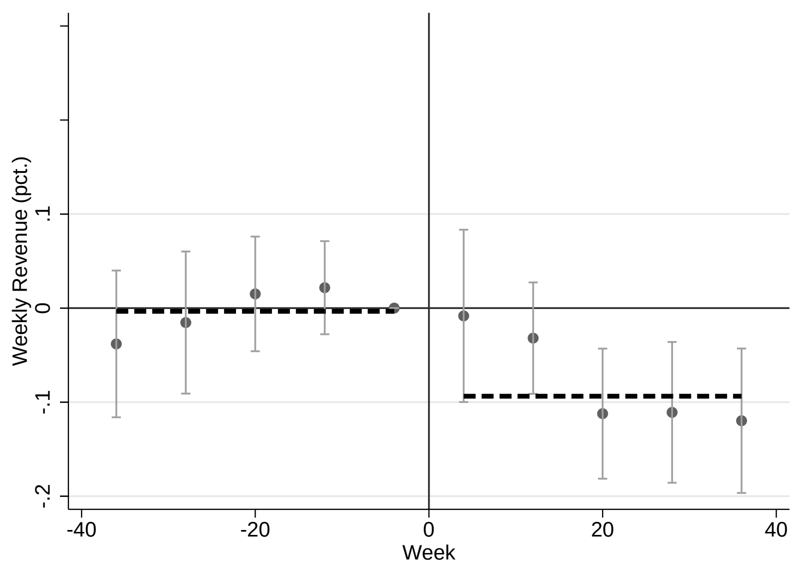
(b) Low-rated Firms

Note: This figure presents event-time estimates from a version of equation 1.5. The dependent variables are the inverse hyperbolic sine transformation and should be interpreted as percentage changes. The unit of observation is business-week. Panel A (B) includes only businesses with rating above the 75th percentile (below the 25th percentile) in the city. The dots represent point estimates from regressing the dependent variable on a treatment indicator interacted with nine-week bins, and business and week-state fixed effects. The treatment indicator compares cities that experienced almost no change in the percentage of businesses available on the platform to cities that experienced meaningful changes. The coefficient in the first period prior to implementation is normalized to zero. The vertical bar represent 95% confidence intervals, where standard errors are clustered at the city level.

Figure 6: Impact of Entry on Weekly Revenue Per Business by Firm Quality (Percentages)



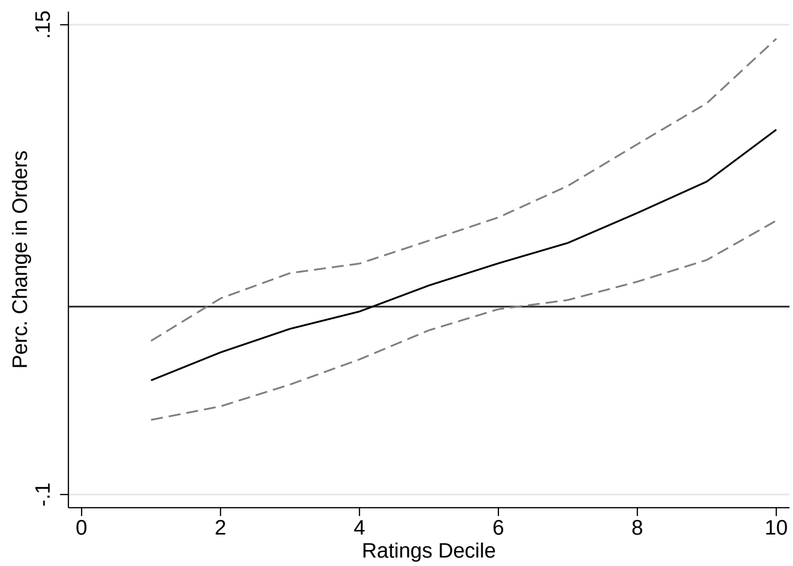
(a) High-rated Firms



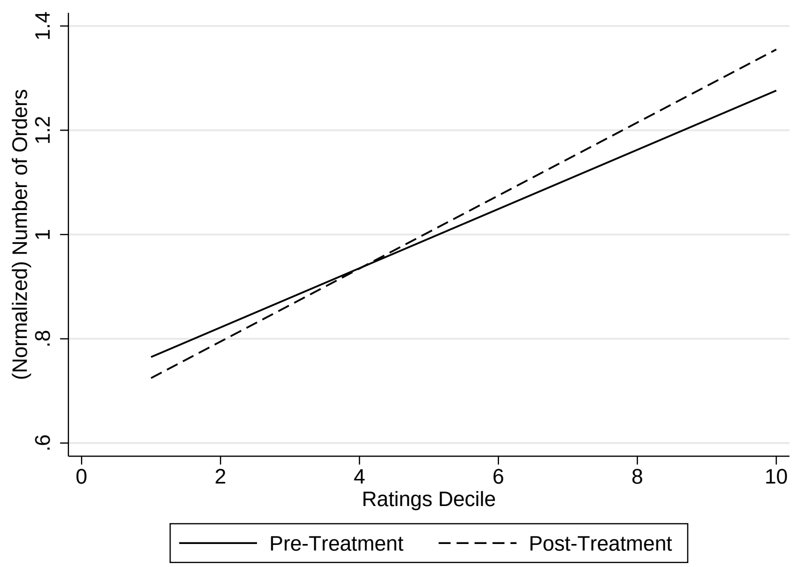
(b) Low-rated Firms

Note: This figure presents event-time estimates from a version of equation 1.5. The dependent variables are the inverse hyperbolic sine transformation and should be interpreted as percentage changes. The unit of observation is business-week. Panel A (B) includes only businesses with ratings above the 75th percentile (below the 25th percentile) in the city. The dots represent point estimates from regressing the dependent variable on a treatment indicator interacted with nine-week bins, and business and week-state fixed effects. The treatment indicator compares cities that experienced almost no change in the percentage of businesses available on the platform to cities that experienced meaningful changes. The coefficient in the first period prior to implementation is normalized to zero. The vertical bar represent 95% confidence intervals, where standard errors are clustered at the city level.

Figure 7: The Impact of Entry on Return to Ratings in Terms of Weekly Number of Order



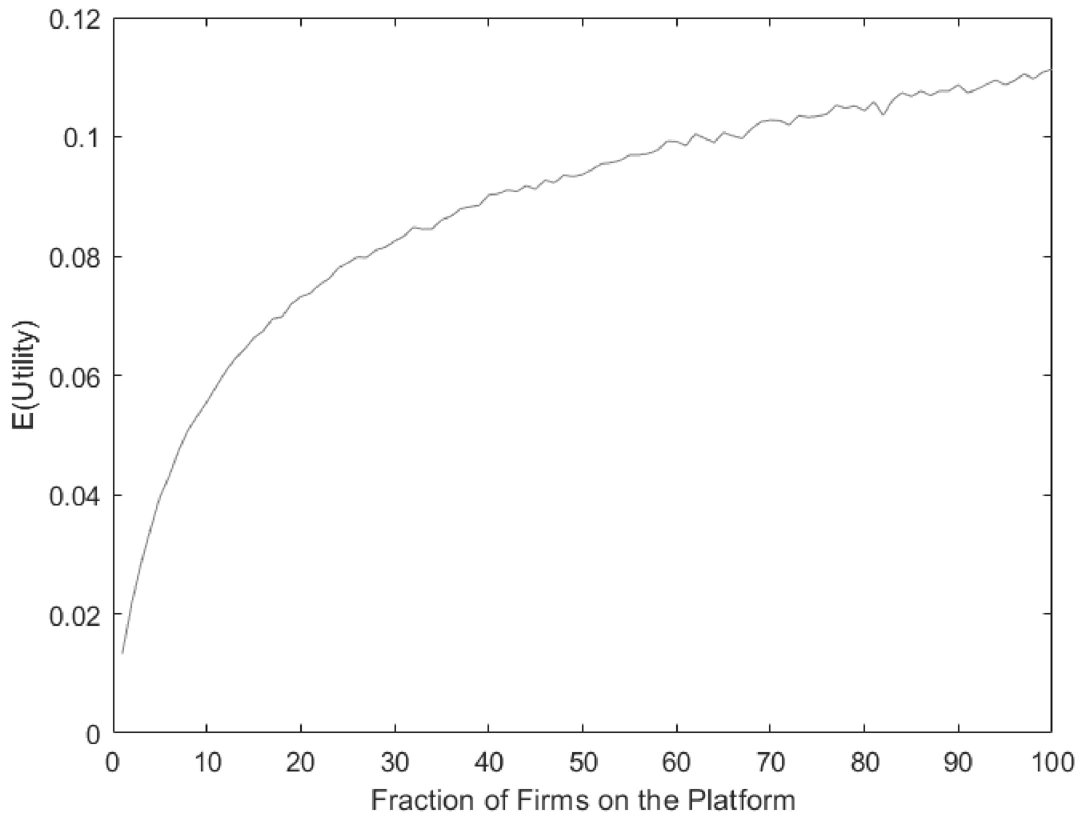
(a) Effect by Rating Decile



(b) Return to Ratings

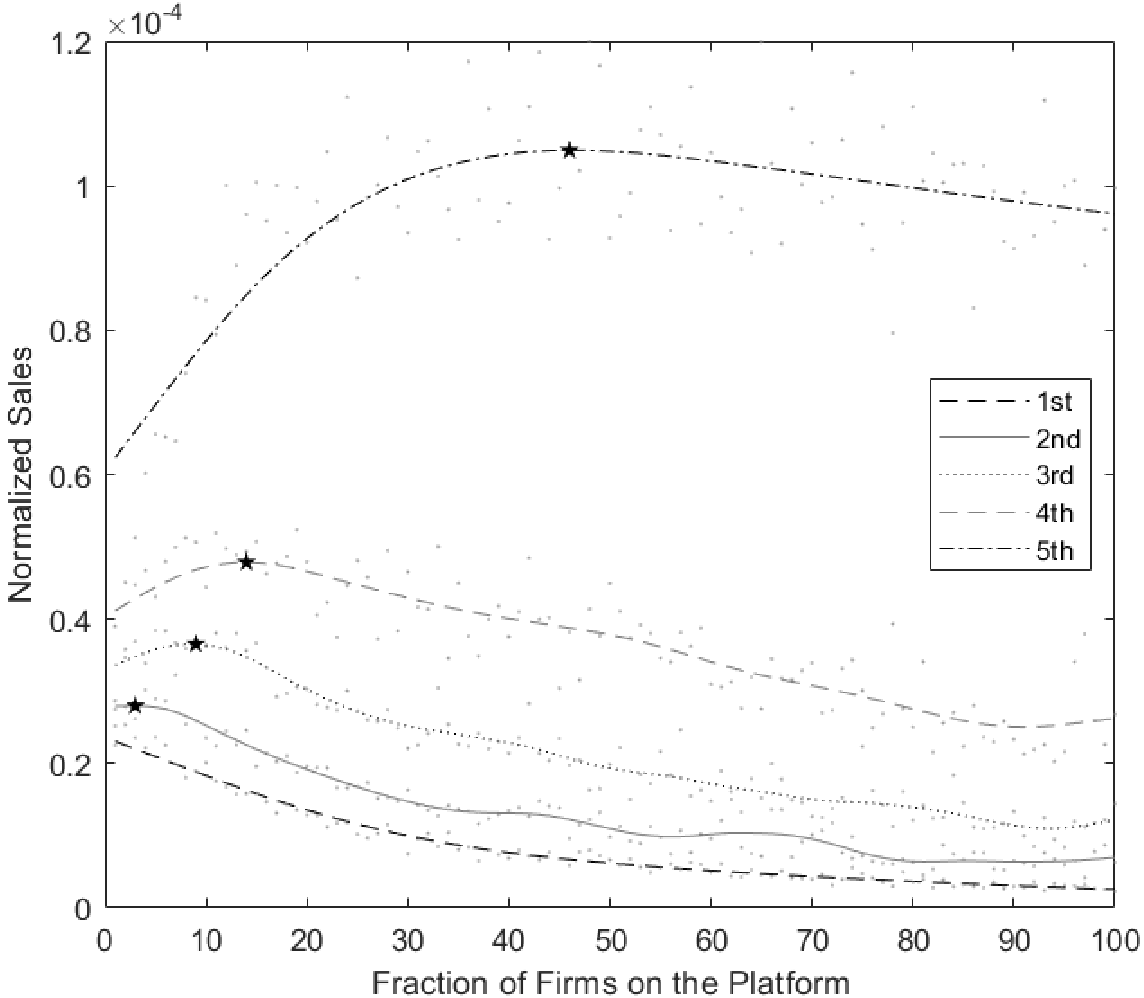
Note: This figure presents the change in correlation between rating decile within city and the percentage change in weekly revenue. The unit of observation is business-week. The solid line in Panel A presents the point estimates of triple interactions between treatment status (sharp treatment definition), an indicator for post integration, and rating decile dummies. The regression includes business and week-state fixed effects. The dashed lines represent 95% confidence intervals, where standard errors are clustered at the city level. Panel B presents the relations between rating deciles and revenue before and after implementation, in treated cities.

Figure 8: Expected Utility by Fraction of Firms on the Platform



Note: This figure present the change in expect utility as the share of firms on the platform grows. It depicts the average results of 500,000 simulated markets. The parameters used to simulate the data are presented in column 3 of Table 12. The simulation algorithm is described in appendix A.5.

Figure 9: Simulation Results of Average Per-Firm Sales by Platform Size and Rating Quintile



Note: This figure present the change in number of sales by rating quintile as the share of firms on the platform grows. The parameters used to simulate the data are presented in column 3 of Table 12. The simulation algorithm is described in appendix A.5. Each gray dot represent the average over one thousand simulations, and the dashed and solid line are the smoothing spline by rating quintile. The stars mark the maximum of each smoothing spline.

Tables

Table 1: Descriptive Statistics by

	Mean	SD	Min	Max
A. Businesses (N=56,493)				
Ratings	3.63	0.61	1	5
Dollar Ratings	1.61	0.51	1	4
Weeks on YTP	94.8	23.8	1	114
Fraction on YTP (pre)	0.051	0.026	0	0.4
Share Change (Business-Level)	0.020	0.013	0	0.2
B. Cities (N=3,965)				
Total Businesses	424.3	1448.9	11	42180
Total on YTP (pre)	14.2	85.9	1	3459
Fraction on YTP (pre)	0.032	0.028	0	0.4
Share Change (City-Level)	0.012	0.018	0	0.2

Note: Panels A and B of this table report the characteristics at the business and city levels, respectively.

Table 2: The Effect of Entry on Market-Level Outcomes

	(1)	(2)	(3)
	Panel A: Weekly Unique Users		
Treat*Post	0.364***	0.482***	10.271***
	(0.015)	(0.020)	(0.634)
	Panel B: Weekly Orders		
Treat*Post	0.367***	0.486***	10.343***
	(0.015)	(0.020)	(0.639)
	Panel C: Weekly Revenue		
Treat*Post	0.587***	0.762***	17.226***
	(0.032)	(0.042)	(1.202)
Observations	327993	226157	327993
# of Clusters	3964	2781	3964
Treatment Def.	Median	25<>75	Change

Note: This table reports regression coefficients from nine separate regressions, three per panel. An observation is city-week, including both incumbent and newly added businesses. The dependent variables are the inverse hyperbolic sine transformation of the outcomes indicated in sub-headings and should be interpreted as percentage changes. Treatment status definitions are indicated below the table and are described further in the text. All regressions include city and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: The Effect of Entry on Firms-Level Number of Weekly Orders and Revenue

	(1)	(2)	(3)
	Panel A: Weekly Orders		
Treat*Post	0.007	0.004	0.388*
	(0.005)	(0.007)	(0.161)
	Panel B: Weekly Revenue		
Treat*Post	0.042***	0.044**	2.068***
	(0.014)	(0.018)	(0.416)
Observations	4409516	2623347	4409516
# of Clusters	3964	2781	3964
Treatment Def.	Median	25<>75	Change

Note: This table reports regression coefficients from six separate regressions, three per panel. An observation is business-week. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders (Panel A) and weekly-revenue (Panel B), and should be interpreted as percentage changes. Treatment status definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: The Effect of Entry on Incumbent Firms by Firm Quality

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Weekly Number of Orders					
Treat*Post	0.036***	0.050***	0.042***	0.065***	1.106***	1.432***
	(0.007)	(0.010)	(0.009)	(0.012)	(0.202)	(0.269)
Treat*Post*Low	-0.062***	-0.098***	-0.084***	-0.119***	-1.654***	-2.342***
	(0.010)	(0.013)	(0.012)	(0.015)	(0.286)	(0.377)
$\beta 1 + \beta 2$	-0.026	-0.048	-0.042	-0.054	-0.548	-0.910
Pvalue	0.001	0.000	0.000	0.000	0.015	0.000
	Panel B: Weekly Revenue					
Treat*Post	0.098***	0.119***	0.121***	0.158***	3.793***	4.059***
	(0.017)	(0.025)	(0.022)	(0.031)	(0.508)	(0.665)
Treat*Post*Low	-0.121***	-0.201***	-0.173***	-0.250***	-4.016***	-5.639***
	(0.026)	(0.033)	(0.032)	(0.040)	(0.679)	(0.893)
Observations	4409516	2173244	2623347	1321619	4409516	2173244
# of Clusters	3964	3875	2781	2725	3964	3875
$\beta 1 + \beta 2$	-0.022	-0.082	-0.052	-0.092	-0.223	-1.580
Pvalue	0.279	0.000	0.049	0.000	0.691	0.015
Treatment Def.	Median	Median	25<>75	25<>75	Change	Change
Quality Def.	Median	25<>75	Median	25<>75	Median	25<>75

Note: This table reports regression coefficients from twelve separate regressions, six per panel. An observation is business-week. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders (Panel A) and weekly-revenue (Panel B), and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding P value. The interaction between post and quality-level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: The Effect of Entry on Incumbent Firms' Item-Level Prices

	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.044*	-0.014	0.052	0.024	-0.007	-0.002
	(0.025)	(0.027)	(0.034)	(0.040)	(0.033)	(0.048)
Treat*Post*Low			-0.028	-0.033	-0.025	-0.074
			(0.056)	(0.085)	(0.054)	(0.083)
Observations	5690994	2803318	5690994	2445849	2803318	1271390
# of Clusters	1703	1009	1703	1356	1009	804
$\beta 1 + \beta 2$			0.024	-0.008	-0.032	-0.076
Pvalue			0.523	0.906	0.464	0.277
Treatment Def.	Median	25<>75	Median	Median	25<>75	25<>75
Quality Def.			Median	25<>75	Median	25<>75

Note: This table reports regression coefficients from six separate regressions. An observation is item-business-week. The sample includes only prices for the most popular items, and excludes item modifications. See text for details. The dependent variable is the inverse hyperbolic sine transformation of item price and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding P value. The interaction between post and quality-level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include item-business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: The Effect of Entry on Incumbent Firms' Subsequent Yelp Ratings

	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.006**	0.008*	0.007**	0.007*	0.009**	0.012**
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
Treat*Post*Low			-0.005	-0.005	-0.004	0.003
			(0.006)	(0.008)	(0.007)	(0.010)
Observations	1449437	850538	1449437	687184	850538	408278
# of Clusters	3562	2420	3562	3380	2420	2314
$\beta 1 + \beta 2$			0.003	0.001	0.005	0.015
Pvalue			0.598	0.864	0.502	0.078
Treatment Def.	Median	25<>75	Median	Median	25<>75	25<>75
Quality Def.			Median	25<>75	Median	25<>75

Note: This table reports regression coefficients from six separate regressions. An observation is business-week. The dependent variable is the inverse hyperbolic sine transformation of the average rating received in a given week and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding p-value. The interaction between post and quality-level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: The Effect of Entry on Incumbent Firms' Advertising Purchases

	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	-0.027*	-0.035**	-0.034*	-0.034	-0.047**	-0.081**
	(0.014)	(0.018)	(0.019)	(0.027)	(0.024)	(0.035)
Treat*Post*Low			0.011	0.050	0.016	0.096**
			(0.026)	(0.033)	(0.033)	(0.042)
Observations	4409516	2623347	4409516	2173244	2623347	1321619
# of Clusters	3964	2781	3964	3875	2781	2725
$\beta1 + \beta2$			-0.023	0.016	-0.031	0.015
Pvalue			0.250	0.419	0.208	0.540
Treatment Def.	Median	25<>75	Median	Median	25<>75	25<>75
Quality Def.			Median	25<>75	Median	25<>75

Note: This table reports regression coefficients from six separate regressions. An observation is business-week. The dependent variable is the inverse hyperbolic sine transformation of total advertising expenditure on Yelp and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding P value. The interaction between post and quality-level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Heterogeneity in the Effect of Entry on Incumbents' Outcomes by Differentiation Between Incumbents and Entrants

	Num. of Orders (Prc.)		Revenue (Prc.)	
	(1)	(2)	(3)	(4)
Share Same Type*Post	-0.010*	-0.029**	-0.002	-0.064*
	(0.006)	(0.015)	(0.014)	(0.037)
Observations	3223311	3223311	3223311	3223311
# of Clusters	1917	1917	1917	1917
Similarity Definition	Positive Change	Continuous	Positive Change	Continuous

Note: This table reports regression coefficients from four separate regressions. An observation is business-week. The sample includes only cities that received above-median change in the share of businesses on YTP. The dependent variables are the inverse hyperbolic sine transformation of weekly number of orders and weekly revenue, and should be interpreted as percentage changes. Coefficients represent the interaction between the measure of similarity, and a dummy for post implementation and treatment status. In columns (1) and (2), the measure is an indicator for whether any business in the same category were added, whereas in columns (3) and (4), the measure is the share of business of the same food category as the incumbent out of the total number of added business. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: The Effect of Entry on Consumers' Search Behavior on YTP

	(1)	(2)	(3)
	Panel A: Number of Searches		
Treat*Post	0.011	0.024**	0.374*
	(0.006)	(0.008)	(0.162)
	Panel B: Number of Views		
Treat*Post	-0.014	-0.006	0.040
	(0.009)	(0.012)	(0.236)
	Panel C: Session Duration		
Treat*Post	0.355***	0.503***	8.016***
	(0.023)	(0.028)	(0.782)
	Panel D: Levenshtein Distance		
Treat*Post	-0.033***	-0.049***	-0.905**
	(0.012)	(0.017)	(0.354)
	Panel E: Generic Queries		
Treat*Post	0.000	-0.000	-0.041
	(0.003)	(0.004)	(0.084)
Observations	159912	94398	159912
# of Clusters	3729	2548	3729
Treatment Def.	Median	25<>75	Change

Note: This table reports regression coefficients from fifteen separate regressions, three per panel. An observation is business-week. The sample includes only Yelp sessions that ended in an order on YTP. The dependent variables are the per-business inverse hyperbolic sine transformations, and should be interpreted as percentage changes. Outcomes are indicated in the sub-headers and described further in the text. The interaction between post and quality-level indicators is omitted for brevity. Treatment status definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10: Testing for the Importance of Ordering of Search Results

	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue
Treat*Post	0.049***	0.064***	0.112***	0.150***
	(0.010)	(0.013)	(0.026)	(0.031)
Treat*Post*Low	-0.103***	-0.126***	-0.215***	-0.271***
	(0.013)	(0.016)	(0.034)	(0.041)
Observations	2098250	1274303	2098250	1274303
# of Clusters	3830	2685	3830	2685
$\beta 1 + \beta 2$	-0.054	-0.063	-0.102	-0.120
Pvalue	0.000	0.000	0.000	0.000
Treatment Def.	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from four separate regressions. An observation is business-week. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding p-value. The interaction between post and quality-level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects, as well as the average search results rank for a business in a given week. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 11: Structural Model Parameter Estimates

	(1)	(2)	(3)	(4)
	Panel A: Utility Function			
Mean Effects				
Dollar Rating	-0.015	-1.554***	-2.393***	0.648
	(0.011)	(0.215)	(0.292)	(.874)
Rating	0.224***	0.265***	0.289***	0.311***
	(0.016)	(0.018)	(0.020)	(0.041)
Ranking (quantile)	0.020***	0.048***	0.029**	-0.043
	(0.007)	(0.009)	(0.013)	(0.029)
Ranking in Category	-0.089***	-0.092***	-0.080***	-0.051**
	(0.001)	(0.001)	(0.004)	(0.022)
Average Income in Zip-Code	-0.081*	-0.809***	-0.598***	0.216
	(0.043)	(0.111)	(0.149)	(0.219)
Random Coefficient (Dollar Rating)				
σ_v			1.361***	1.772
			(0.302)	(2.130)
Income \$50,000-\$100,000				-1.032
				(1.694)
Income above \$100,000				2.724
				(5.483)
Female				-2.104
				(5.619)
Age 25 to 44				-4.138
				(18.72)
Age above 45				-3.991
				(18.58)
Food Category FE	Yes	Yes	Yes	Yes
Instrument	No	Yes	Yes	Yes
	Panel B: Entry Costs			
Average μ	.071 ***	.205***	.197***	.203***
	(0.001)	(0.006)	(0.10)	(0.009)
σ	0.010 ***	0.091***	.093***	0.098***
	(0.003)	(0.003)	(0.005)	(0.005)
Heterogeneous μ	No	No	Yes	Yes
Observations	124061	124061	124061	124061
# of markets	5500	5500	5500	5500

Note: This table reports four separate estimates of versions of the model described in Section 1.6. Panel A presents the parameters of the utility function, and Panel B presents the parameters of the entry cost distribution. Columns (1) and (2) present the model without user heterogeneity. Columns (3) and (4) allow for the coefficient of dollar rating to vary by unobserved- and unobserved-and-observed user characteristics, respectively. Columns (3) and (4) also allow the distribution of entry costs to vary by household income-levels. For these columns, only the weighted average μ over income groups is presented. All regressions include CSA fixed effects. Robust standard errors are in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 12: Simulations of Welfare, Market Size, and Firm-Level Outcomes

	10%	20%	40%	60%	80%	100%
Δ Welfare	0.40	0.85	1.28	1.51	1.64	1.81
Δ Market Size	0.41	0.94	1.58	1.98	2.19	2.58
Δ Sales By Rating Quantile						
1st Quantile	-0.13	-0.36	-0.64	-0.76	-0.83	-0.88
2nd Quantile	-0.12	-0.34	-0.55	-0.64	-0.79	-0.75
3rd Quantile	0.02	-0.17	-0.37	-0.53	-0.62	-0.67
4th Quantile	0.10	0.03	-0.12	-0.26	-0.38	-0.38
5th Quantile	0.12	0.32	0.47	0.46	0.40	0.36

Note: This table reports the average results of 500,000 simulated markets. The parameters used to simulate the data are presented in Column (3) of Table 12. The simulation algorithm is described in appendix A.5. The tables presents the percentage change in outcomes from the baseline. In the baseline, 5% of firms in the market are on the platform. Outcomes are indicated in row names and the subheading. Column headers indicate the simulated share of firms on the platform.

Chapter 2

The Impact of Prices on Firm Reputation

Chapter abstract: We explore the impact of prices on a firm's reputation. We analyze data on prices, orders, and ratings from a large online review and ordering platform. Looking at narrow windows around the timing of menu price changes, we find that online reviews are influenced by price changes and that increasing prices tends to harm a firm's reputation; an increase of 1% in item price leads to a decrease of up to 5% in the ratings left by users. Consistent with this, the distribution of ratings for cheaper restaurants is similar to that of more expensive restaurants. Finally, these effects do not seem to be driven by consumer retaliation against price changes, but rather by changes in absolute price levels.

2.1 Introduction

A large literature has explored the impact of reputation on firm outcomes. A better reputation can allow firms to increase their prices, demand, and profitability (Chevalier and Mayzlin, 2006, Cabral and Hortacsu, 2010). Accordingly, firms also endogenously make efforts to affect their reputations depending on market conditions (Jin and Leslie, 2009, Board and Meyer-ter Vehn, 2013). At the same time, while reputation affects prices, prices also have the potential to affect reputation. For example, when product quality is not immediately observed by customers, increasing prices might signal higher quality (Shapiro, 1983, Milgrom and Roberts, 1986). On the other hand, higher prices can also lead to a worse reputation if reputation is a function of quality conditional on price.

User generated feedback had emerged as one of the dominant forms of firm reputation. Almost all transaction platforms now have some form of rating mechanism to provide consumers with a signal of sellers' quality (for example, Amazon, Ebay, Airbnb). In online markets, user ratings are often the only sources of seller reputation. In addition, consumer reviews are increasingly impacting firms and product reputation in offline markets as well (IMDB, Yelp, TripAdvisor).

In this paper, we explore the causal impact of price on restaurant reputation, as measured by its online ratings. We restrict our attention to item-level price variation and study the impact of sharp changes in price on feedback received just-before and just-after price changes.

We find that increasing prices has a negative effect on subsequent ratings. In particular, a 1% increase in price leads to a 0.05-0.14 decrease in rating on a scale of 1 to 3, which is approximately, 2.5%-5% decrease for the average feedback. This effect becomes increasingly important when considering the fact that the average price change is about 3%-9%. This result has direct implications to consumers, firms, and platform designers.

To conduct our analysis, we use data from Yelp, the commonly used consumer-review website for local businesses. We begin by looking at data on the cross sectional distribution of the Yelp star-ratings for restaurants. If ratings were simply a proxy for quality, one might expect a positive relation between price and quality - with more expensive restaurants having much higher ratings. In contrast, we find that the distribution of ratings for cheap restaurants is very similar to the distribution of expensive restaurants. For example, in Berkeley, the Michelin restaurant, Chez Panisse, which is considered the birthplace of Californian cuisine, has the same four-star rating as Top Dog, students' favorite hole-in-the-wall hotdog stand. More generally, we find that the rating distribution of the cheaper compared to the most expensive restaurants are surprisingly 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 explore the causal link, we then turn to data from Yelp Transactions Platform (YTP), an online platform for takeout and delivery from local restaurants. We obtain access to item-level information on all food ordering- and delivery-transactions finalized on YTP. To identify the effect of price on ratings, controlling for unobserved quality, we introduce two main specifications: First, by including a myriad of fixed effects in the regression specification in order to account for item-level and time variation, and second by comparing transaction-level reviews just-before and just-after sharp price changes. In our preferred specification, we find that a 1% increase in price leads to a *decrease* of about 0.11 points on a scale of 1 to 3, about 4%, in subsequent ratings. 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.

While we cannot say whether some firms would have been better off by not increasing their prices, our results do point to a tradeoff - price increases come with a reputational cost that leaves firms worse off, relative to a world in which ratings do not adjust with price. In addition to changing firms' incentives, this dynamic has material implications to the value of the rating to consumers. If consumers are unable to unpack the impact of historic prices on rating, then this would create a wedge between items' true quality and the perceived quality (or reputation).[1] Platform makers can alleviate this concern and improve existing reputation mechanisms by redesigning the rating mechanism to account for the impact of historic prices on reviews received.

One potential mechanism is that prices are serving as a signal of quality for consumers who are then disappointed after eating an expensive but mediocre piece of pizza. If this is the case, we should expect the effect to be larger among users who have not previously ordered

[1] In principle, consumers could potentially infer the true quality if they know the history of prices and the data generating process. However, in practice, most platforms do not present historical prices, Thus, it would be difficult for a customer to know a business's entire pricing history, and back out the true quality.

from that business. Consistent with this, we find that the effect is larger and generally more statistically significant for people who are ordering an item for the first time at a restaurant relative to people who have ordered before. This is suggestive evidence that consumers respond to the increase in price level rather than retaliate against the firm and using low ratings as punishment for raising prices.

Overall, our results contribute to the literature on firm quality and reputation. Existing work has shown that high ratings allow firms to increase price and sales (Livingston, 2005, Jin and Kato, 2006, Chevalier and Mayzlin, 2006, Resnick et al., 2006, Cabral and Hortacsu, 2010, Luca, 2016). Our work examines the feedback loop in the other direction, focusing on the impact of price on reputation. We thus contribute more generally 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). and the benefits of introductory pricing (Cabral et al., 1999, Schlee, 2001).

Our findings also contribute to the literature on reputation systems. User ratings can help to mitigate asymmetric information between buyers and sellers, especially in online settings (Dellarocas, 2003, Bar-Isaac et al., 2008, Tadelis, 2016). 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). We show that firms can also use prices as a lever to impact future ratings and improve (or worsen) reputation.

Finally, while ratings are often thought of as a proxy for quality, a growing literature documents biases and issues with consumers' reviewing process (De Langhe et al., 2016, Nosko and Tadelis, 2018, Horton and Golden, 2018, Fradkin et al., 2019). Our work helps to shed light on the content of reviews, showing that ratings take into account both quality and price. These new insights contribute to the literature on the design of rating systems (Bolton et al., 2013, Li and Xiao, 2014, Dai et al., 2018).

2.2 Data and Empirical Design

2.2.1 Settings and Data

We study the impact of price on user generated reviews in a portion of the food delivery-service industry in the United States, a 35 billion dollar industry that is expected to grow at an average annual rate of more than 20% in the next 10 years.[2] The main focus is the Yelp Transactions Platform, an online platform launched in 2013 by the consumers' review website, Yelp. YTP enables users to order food delivery and pickup from local restaurants through several food-delivery services.

YTP operates as a part of the standard Yelp website and features a subset of restaurants available on Yelp. Shoppers are automatically directed to the platform by applying the "Delivery" or "Takeout" filters, or by using similar words in a search query.[3] Figure 1

[2]According to a UBS Investment Bank report.

[3]There are alternative ways to access YTP: First, shoppers can start an order directly from the business page. Second, when a user performs a search on the standard Yelp website, businesses that are YTP affiliates

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 the specific menu items they are interested in and finalize the transaction.

Following a transaction on YTP, users are prompted with the option to leave a review for the restaurant (Figure 1c).[4] Feedback prompt may appear on the Yelp application or via email, depending on the method by which the order was carried out. Response rates are relatively high, approximately 20% of users leave a review. 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). Finally, if the consumer had a bad experience, she is solicited to list the specific issues for the unfavorable experience.

We collaborate with Yelp to gain access to proprietary transaction-level data. Establishment-level 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 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 more detailed discussion can be found in Appendix B.2. The data includes all food orders completed on YTP during that period.[5] 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. Unfortunately, as presented in Figure A3, much of the price variation we find is due to spurious price changes. Though we cannot always pin down the source for these spurious price changes, anecdotal and interviews with Yelp employees suggest that these are the results of item modification. For instance, consumers may add a topping to a pizza, ask for extra rice, or substitute chicken with shrimp. These changes are often associated with additional costs. We generally observe item-level pricing and description, which sometimes detail whether the item includes an add-on. The fixed effect regressions, which are used in

will have an "Order Now" button next to their name on the search results.

[4]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.

[5]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, revenue, or the precise number of users.

the second part of the analysis, exclude items with ‘suspicious’ descriptions.

Nevertheless, we identify multiple cases where order modifications are not documented in the data. To address this issue, we develop a simple algorithm that further restricts our sample to “true” price changes. We use the algorithm when focusing on reviews received right around price changes. The specific details of the sample selection algorithm are presented in Appendix B.2. Intuitively, we begin by restricting our attention only to items with sufficient observations and without any document price modifications. Then, we remove prices that appear infrequently in the data or prices that do not appear sequentially. Finally, we exclude periods in which we observe more than one prevailing price.[6]

2.2.2 Research Design and Empirical Specification

We are interested in identifying the impact of prices on subsequent ratings. Using prices as a predictor is always problematic since prices are not randomly assigned, but rather strategically determined by the firm. Thus, we expect prices to be correlated with unobserved characteristics such as input costs, changes in quality, or shifts in demand. The main identifying assumption is that the impact of unobservable factors diminishes when looking at the same item within a small time frame.

In the lone enough run, we can expect unobserved demand shocks to impact price. For example, restaurants may become more popular and neighborhoods may become trendier. These secular trends however, tend to occur over a long time horizon and are not expected to vary at the weekly-level. In fact, we find 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 the menu items ever change prices. The most plausible explanation is that restaurants are characterized by high menu costs (Hobijn et al., 2006, Bils and Klenow, 2004). Thus, even if there are sharp changes impacting a restaurant’s demand or cost, such a good review in a famous journal or input shortage, menu costs create a friction that hinders updating prices in the very short run.

Moreover, 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 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.

Finally, some changes, such as kitchen renovation or hiring a new chef might impact all dishes simultaneously. Again, it seems unlikely that price would adjust instantly as one can expect a transition period for these improvements to affect food quality. Moreover, economics theory predicts that improvements in quality would lead to higher, rather than lower, prices. Nevertheless, as a robustness test, we estimate an additional specification, which takes ad-

[6] As part of the robustness tests, we test the sensitivity of our result to altering the core assumptions of the algorithm and making it more or less conservative.

vantage 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.

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

$$Y_{jt} = ln(Price_{jt}) + X_{jt} + \gamma_w + \delta_j + \mu_{wj} + \epsilon_{jt} \tag{2.1}$$

An observation is one item in a transaction. w is the index for week, j denotes item-business combination , and t is an index for specific dates. Y denotes the outcome of interest: Rating received following the order, 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. 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.

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

$$Y_{jt} = ln(Price_{jt}) + X_{jt} + \lambda_G + \epsilon_{jt} \tag{2.2}$$

Here—beyond the indices and variables defined in equation 2.1— G is an index for item-business X price change and λ is the fixed effect per item price change. 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 2.2.1 and Appendix B.2.

2.3 Results

2.3.1 Motivating Evidence

To motivate the main analysis we start by analyzing the relation between prices and ratings in the standard Yelp website. Particularly, we examine the relation between the Yelp Star Rating and the Yelp Dollar Rating. 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.

The main results are presented graphically in Figure 2. 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, sushi, etc. The figure suggests that, though higher prices are correlated with higher ratings, the impact 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),[7] even though, on average, they are more than 4 times more expensive.

This suggests that more expensive restaurants do not receive overwhelmingly higher ratings, consistent with our hypothesis that ratings are price adjusted prices. 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 any compensation for the hefty price tag. Nevertheless, in the next subsection we attempt to more directly identify the relation between prices and ratings.

2.3.2 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 and specific items. In particular, we use price variation within items to estimate the impacts 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 2.1, 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 with the least restrictive set of fixed effects, without any item- or week-level controls. 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. Recall that the average price change is about 3%, which implies an increase of about 0.018 in rating, or about 0.006% change. This positive but weak result ecos the pattern presented in Figure 1.

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 impact of price on the (linear) probability of receiving the highest rating (Great) or not receiving the lowest rating (Bad), respectively. We find that prices impact 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)).

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

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 our attention only 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 description and discussion on the rationale guiding our decisions is presented in Appendix B.2.

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 included in the sample is substantially lower than present in Table 1. This approach, while reducing our sample size, provides arguably the cleanest estimate of the impact 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 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 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 might be an artifact 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.2.[8] The estimated magnitudes are consistent with the ones presented in Column (1) and remain statistically significant across specifications. Similar to Table 1, 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. While the effects on both margins are statistically significant, these two columns suggest that the effect is driven mostly by reduction in the highest score, with about 66% attributed to reduction in the highest ratings possible 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. In particular, we examine the importance of the allowed window around price changes.[9] Finally, to relax the linear probability assumption, we estimate the

[8]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. Anecdotally, the lower price is usually more prevalent than higher prices. In addition, we usually think about (and observed) product modifications as increasing prices, so the lower price is more likely to be the standard menu 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.

[9]Consistent with Column 1 in Tables 1 and A1, the effect diminishes as the window becomes larger, but

impact of prices non-linearly using ordered logit to re-estimate Column (1), and conditional logit for Columns (4) and (5). Appendix B.3 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.

2.3.3 Mechanism and Alternative Research Designs

The Role of Retaliation In the previous section we found that price has a deleterious effect on firms' ratings. While we cannot disentangle all mechanisms that could explain this finding, here we discuss some potential mechanisms and role out alternative explanations.

One potential explanation is that consumers are not responding to price levels, but rather to price changes, i.e. consumers are displeased with the price increase and "punish" the seller by leaving a negative review. For instance, if the old prices serve as reference points then higher prices may be perceived by consumers as a loss (Tversky and Kahneman, 1979). Alternatively, consumers might be expressing their discontent rather than reducing transactions with the firm (Hirschman, 1970). Whatever the underlying psychological mechanism is, one assumption of these explanations is that consumers are familiar and accustomed to the former prices. Thus, we expect this effect to attenuate for users who haven't ordered from the specific restaurant in the past.

Table 3 shows the impact of price changes on ratings when restricting the sample only to users who order from a specific restaurant for the first time. The table has a similar structure to Table 2, but with the sample restricted to new users. Column (1) presents the estimation results for our main specification. The coefficient on log price is -0.126, which 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.[10]

We interpret these results as suggesting 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. Though, it still might be the case that consumers are, to some extent, familiar with restaurant prices from shopping offline, in an unshown analysis, we restrict the sample only to repeating consumers and find much smaller point estimates. In addition, the point-estimates are not statistically significant, though this might be due to the fact that repeated users are far less likely to leave a review and thus the sample is considerably smaller.

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

becomes larger in magnitude when we narrow the window.

[10] 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.

alternative research designs.

First, as discussed in Section 2.2.1 and presented in Figure A3, for most items there are no price changes and when there are price changes, most of them 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 irregularly 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 our attention only to weeks in which we observe more than 5, 10, or 20 of the restaurant's items changing price in a given week; a sizable change given that the average restaurant in our sample only sells about 30 items. The results are presented in Appendix Table A3. Similar to the main analysis, we find a consistent negative and statistically significant relation between price and ratings. The estimated magnitudes are 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 sold 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: Note that 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.[11] 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 increases from \$7 to \$9 for both partners, but partner B updates the price before partner A does. This means that there is a short time window where we see the same item sold at both \$7 and \$9 for different consumers. Thus, we can treat prices as randomly assigned and estimate the impact of price increases on ratings.

The results are presented in Table A4. 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 effect seems to be driven mostly by drops in the highest rating. This effect, however, is not statistically significant, possibly due to the small number of observations. 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

[11]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.

item does not live up to the expectations formed based on its hefty price tag.

2.4 Discussion

In this paper we study the impact of price on firm reputation, as measured by its online consumer reviews. 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.

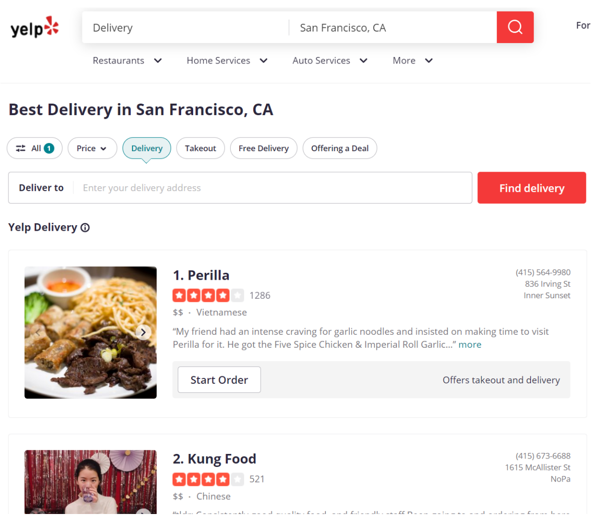
This result has several important implications to consumers, firms, and rating system designers. First, the results speak to the potential welfare gain and value of reputation systems to consumers. For instance, if consumers are unable to unpack the impact of historic prices on ratings, or they have some incorrect beliefs about how raters incorporated prices into their reviews, then this would create a wedge between items' true quality and the perceived quality (or reputation). In particular, this mechanism does more than introducing noise into consumers' decision-making process, but instead have consistent biases in specific directions. Strategic sellers might be tempted to take advantage of misguided consumers to maximize their profit.

Second, consistent with the above analysis, this dynamic might create additional incentives for firms to set low introductory prices. Initial low prices can mechanically boost ratings and allow some firms to later take advantage of their good reputation and increase sales or prices. More generally, our results point to a tradeoff - price increases don't just reduce present demand, but can potentially also reduce future demand by negatively impacting firm reputation. This dynamic might reduce firms' willingness to increase prices, even in response to changes in demand or the cost structure. Finally, platform makers can and improve existing reputation mechanisms by redesigning the rating mechanism to account for the impact of historic prices on reviews received.

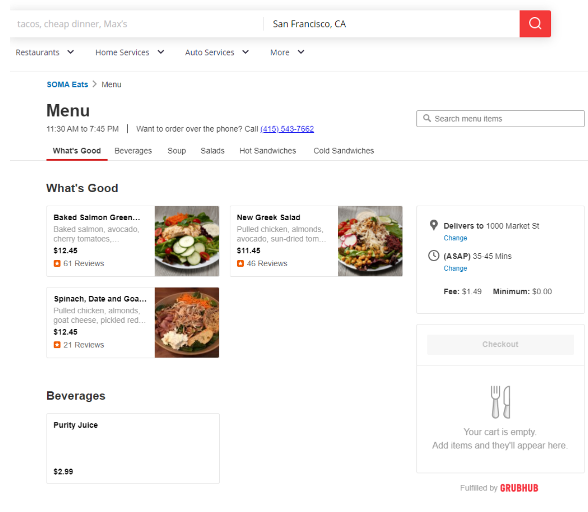
Lastly, in this setting the mechanism driving the adverse impact of prices on ratings is somewhat unclear. We believe the two most plausible explanations are that consumers rate net utility, i.e. quality minus price, rather than quality or that consumer rate the deviation from their ex ante expectation, i.e. value minus expectation, and that expectation is positively correlated with prices. We believe that the answer is a mixture of these two forces: On the one hand, we do find that old 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 do 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 to forces.

Figures

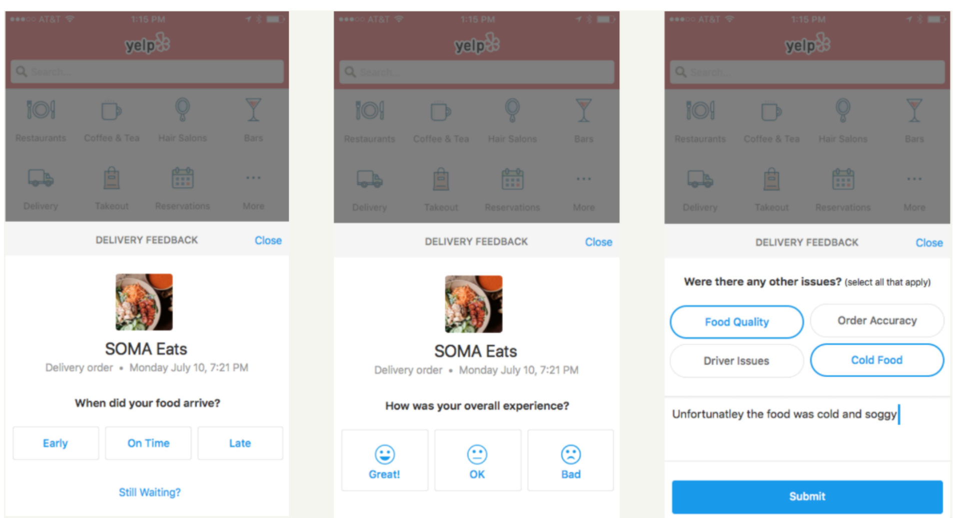
Figure 1: Visualization of the Ordering and Review Process on Yelp Transactions Platform



(a) Searching on YTP



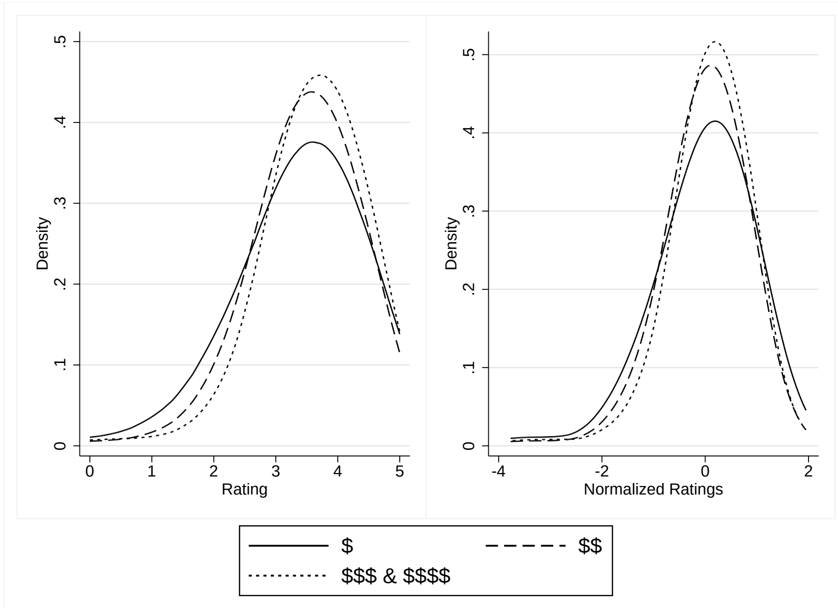
(b) Ordering on YTP



(c) Rating on YTP

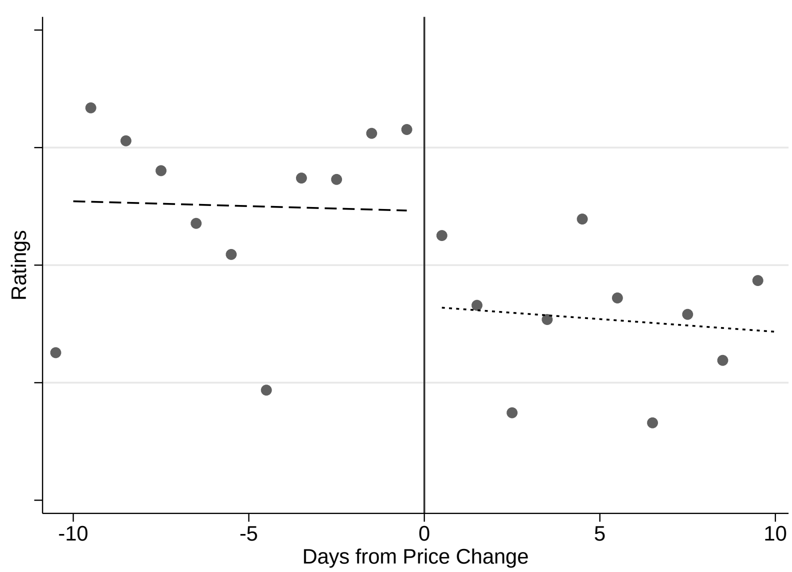
Note: Panel A presents the search results for delivery on YTP around in San Francisco. Panel B presents a menu for a restaurant affiliated, where consumers can pick the specific items and finalize the transaction. Panel C presents review process on YTP.

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



Note: This figure presents the density distribution on Yelp Star-Ratings by Dollar-Ratings in three cities with 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 Impact of Price Change on Ratings



Note: This figure presents the raw relation between ratings and days to price increases. The day of price increase is normalized to zero.

Tables

Table 1: The Impact 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)
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 indicated in the column headers and described further in the text. 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 Impact 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)
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 indicated in the column headers and described further in the text. 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 Impact 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)
Constant	3.064***	3.065***	3.158***	0.986***	1.078***
	(0.102)	(0.101)	(0.174)	(0.086)	(0.041)
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 indicated in the column headers and described further in the text. 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%

Chapter 3

Persistence and the Gender Innovation Gap: Evidence from the U.S. Patent and Trademark Office

Chapter abstract:

In this study, we provide causal evidence that one contributor to the gender "innovation gap" is that women are less likely to persist after an early rejection in the patent process. To provide causal evidence of a persistence channel, we use exogenous variation in the likelihood of early-stage adverse decisions about patentability claims that arises from the random assignment of applications to patent examiners. We find that majority-female innovator teams are less likely than majority-male teams to either appeal or amend applications that receive rejections within the patent prosecution process. Roughly 1/2 of the overall gender gap in awarded patents can be accounted for by the differential propensity of women to exit the application process after a rejection of patent claims at the first stage of the prosecution process (an outcome that is overwhelmingly common, even for applications that ultimately result in awarded patents). We also provide evidence that the gender gap in persistence is reduced when women-led applications have the backing of firms, consistent with a potential role for institutional support in mitigating gender disparities. Gender differences in persistence seem to have little to do with examiner identity.

3.1 Introduction

Innovation plays an important role in the United States' economy and society. Inventors, however, are not representative of the population at large; in 2010, only 18.8% of patents had at least one female inventor, and only 8% of all patents had a woman listed as primary inventor (Milli et al., 2016). More generally, a large literature documents gender differences in STEM performance (Beede et al., 2011, Arcidiacono et al., 2016). In addition, women-led patent teams and even teams including women are significantly more likely to produce female-focused innovations (Koning et al., 2019). Thus, if women are less likely to participate in innovation, then the female population more broadly suffers as a result.

In this paper, we focus on one contributor to the gender innovation gap — gender differences in persistence. In particular, consistent with evidence from other settings (see, e.g., Wasserman, 2018), we argue that women are less likely to follow up on an application after receiving negative initial feedback, which leads to a downstream reduction in patents granted to women. In order to identify the causal effect of persistence, holding constant other channels, we use quasi-exogenous variation in the likelihood of early-stage adverse decisions that arises from the random assignment of applications to patent examiners. Specifically, we use examiner harshness across all *other* applications as a an instrument for initial patent rejection in order to study the heterogeneous responses across female- and male-led innovator teams.

We find that women-led teams are 3-7 percentage points (about 4%-7%) less likely to continue the patent process after an initial rejections compared to men. This differential effect by innovators' gender is magnified when examining whether a patent is ultimately issued; we find that initial negative feedback differentially reduces the probability that a patent is granted by 5.5-10 percentage points (about 8%-14%) more for females compared to their male counterparts. We argue that these results have important implications to understanding the gender innovation gap and prospective public policy to alleviate the gap. Finally, we explore the underlying mechanisms and moderators of gender differentials in persistence.

For our main analysis, we use data from the United States Patent and Trademark Office (USPTO). We use of all patent applications in the United States from 2002 through 2014. The final sample covers over one million applications from US-based teams, both those that received patents and those that did not. The data include basic information on a patent application, including the technology class, the outcome of the application (whether a patent is issued), and innovators full names, which we use to elicit gender. Importantly for our analysis, the USPTO data also include complete prosecution histories, which detail the entire application process. In particular, we can observe each step of the application and communication between the patent examiners and applicants, including rejections, amendments, and appeals.

We examine how men and women respond to rejections within the patenting process. Notably, patent applications are rarely categorically rejected by the USPTO. Rather, they are either implicitly or explicitly abandoned by applicants following what technically are appealable rejections issued by patent examiners (Lemley and Sampat, 2008). Thus, we consider innovator tendency to follow up on an application and amend their claims as a measure of persistence. Naturally, rejections are not randomly assigned and might be potentially correlated with unobservable application attributes, such as patent quality. In order to address this concern, we use the quasi-random assignment of applications to examiners (Similar in spirit to Farre-Mensa et al., 2017, Sampat and Williams, 2019). The main intuition is that examiners differ systematically in their propensity to approve patents at any stage. Lenient examiners are more likely to grant a patent than are stricter examiners, holding the quality of the proposed invention constant. Thus, drawing a harsher examiner increase the probability of initial rejection, regardless of patent quality or other ex ante characteristics.

The analysis yields several insights: To begin with, we find that female-led teams are only marginally more likely to receive initial rejections compared to men (less than 1 percentage point), but are substantially less likely to ultimately receive a patent (over 5 percentage points). We interpret this as suggestive evidence that differential in gender persistence, rather than examiner discrimination, are driving the difference in patents granted. Accordingly, in our main results, we use examiner harshness as an instrumental variable and find that women are significantly less likely to persist in the patent process if they receive an initial rejection, compared to men. For instance, when restricting our attention only to applications filed by individuals, our estimate suggests that an exogenous rejection reduces the percentage of women applicants by 4.3 percentage points (about 5%), compared to men. The effect remains statistically significant when we examine the effect of the proportion of women inventors, or use indicators for whether the innovators team consists mostly or solely of female innovators.

Similarly, we find that initial rejections have stronger downstream implications for female- compared to male-led teams. For instance, when comparing applications filed by mostly-female teams to mostly-male, the estimated effect is 7 percentage points. This implies that women-led teams are, on average, 10% less likely to have a patent granted following an initial rejection, compare to men. In subsequent analysis, we examine whether teams working as part of a firm behave differently than individual inventors. We find that firm-backed applications are considerably more likely to proceed beyond initial rejection. This effect, is more pronounced for female- compared to male-led teams (an additional increase of 3.8 p.p. for female-led teams). We thus conclude that firm backing partially offsets the persistence gap.

Our findings contribute to literature studying the gender innovation gap. Understanding where in the process of innovation women fall out and why this happens is essential in order to develop solutions that address the gender gap in innovation (Jensen et al., 2018, Stewart Stute, 2019, Sarada et al., 2019). More generally, this study speaks to the growing literature on gender differences (Niederle and Vesterlund, 2007, Alan et al., 2020), especially in STEM performance (Griffith, 2010, Arcidiacono et al., 2016). Finally, our results speak to a growing literature studying gender differences in persistence in other settings, such as the work place (Brands and Fernandez-Mateo, 2017), politics (Wasserman, 2018), and crowdfunding (Kuppuswamy and Mollick, 2016).

3.2 Empirical Framework

3.2.1 Data

To study persistence differences between men and women in the patent prosecution process, we use individual patent application data over 12 years from the Patent Office's Patent Application Information Retrieval (PAIR) database. The PAIR data cover all utility patent applications that were filed on or after March 2001. We focus on non-provisional utility patent applications, which are formal applications for new processes, machines, and manufacturing systems. Utility applications are the most common patents issued by the USPTO and make up over 90% of all patents issued annually (USPTO 2012). Because we infer patent applications that are abandoned implicitly based on non-receipt of a granted patent, we focus on applications that reached a final disposition by July 2012 (thus excluding ongoing applications). In addition, we restrict out attention to applications in which all of the team members are United States nationals. The PTO application records include basic information on a patent application, including the technology class, Art Unit, the outcome of the application (whether a patent is issued). Importantly for our analysis, the PTO data also include complete prosecution histories, including the examiner assigned to review an application — which we need to implement our identification strategy. The PAIR data also provides data on the timing of application, each examiner rejection, applicant amendment/appeal, and final patent allowances (including when an application is amended, or resubmitted by an inventor/team).

Note that the gender of the applicant is not an explicit field within the USPTO data. We thus follow a an increasingly common process for imputing the gender of an applicant that relies on publicly available data on the gender distributions of first names, in our case from the US Social Security Administration (SSA) to identify the frequency with which specific names are given to males and females (Jensen et al., 2018). There are 97,310 unique names in the SSA data, for which we construct a gender distribution by name. For example, if there are 9,000 people with the name Carol who were women and 1,000 male Carols, then the name Carol would receive a female proportion of 0.9. We match these names to the 271,000 unique patent applicant names from the USPTO data to assign gender to patent applicants. We use an 80% cutoff threshold and and drop applications for which any inventors' names are assigned either male or female less than 80% of the time. We only include applications for which we can assign gender to *all* inventors on the team. There are about one million applications that constitute the core of this analysis.

The majority of patents are submitted by teams of inventors. We focus on a few different measures of gender composition: (1) whether an application team is composed of 50% or more women (half female), (2) whether it is composed of all women, and (3) the proportion of women on an application. We also conduct a separate analysis in which we restrict attention to applications submitted by individuals. Table 1 provides summary statistics for our sample. Table 2 shows the likelihood of submitting initial and final amendments as well as receiving a patent by composition of the inventor team. Applications from inventor teams including women are less likely to do any of these things than the mean application in our sample. In the next section, we detail our strategy to isolate if gender differentials in persistence may be causing this gap.

We also use other information collected by the USPTO to test for heterogeneity and examine mechanisms underlying our results. For example, using the PTO's assignment data, we can tell whether an inventor is connected to a firm or is applying solo.[1]

3.2.2 Empirical Strategy

The goal of our study is to cleanly identify heterogeneous responses across men and women to negative decisions made by patent examiners, particularly at the First Office Action (FOA) stage of the patent application process. In other words, can we identify gender differences in the likelihood that an applicant "persists" after having initial patent claims rejected (an *initial rejection*). We operationalize "persistence" in this setting as the likelihood of continuing with the patent prosecution process, conditional on rejection at the first stage. While a fraction of patents will be granted based on the original claims, many more will

[1]When an inventor or team of inventors apply for a patent, they are presumed to be the owner of the patent. Applicants can assign their idea to an organization or entity, generally the company for which they worked when they produced the idea.

involve adjustment to an application. In other words, after receiving an initial rejection (the modal outcome in terms of an examiner's initial decision), are women less likely to submit an amended patent application or appeal as well as ultimately receive a patent? The ideal experiment to identify gender-specific responses to patent denials would be to randomize patent denials at the first instance across men and women, thus ensuring that successful applicants and their inventions do not differ systematically from unsuccessful ones ex ante. Gender differences at this stage would then indicate differential selection out of the patent prosecution process.

We model the relationship between receiving an initial rejection and continuing the patent prosecution process (including final receipt of a patent) as follows:

$$Y_a = \beta_1 Initial\ Rej_a + \beta_2 Female_a + \beta_3[Female\ \times\ Initial\ Rej]_a + \mu_{ut} + \epsilon_a \tag{3.1}$$

In this setup, a indexes a patent application, and ut the patent art unit - cross - application year. Y_a is the outcome of interest, usually whether inventors continued the application or whether the application was approved. $Initial\ Rej$ is a dummy for whether the application received an initial rejection, and $Female$ is an indicator for the prevalence of females in the inventors team, as described in Section 3.2.1. We are interested in identifying the coefficient estimate for β_3, which tells us how likely women are to either amend an application or finally obtain a patent, conditional on receiving a rejection at phase one of the examination process, relative to men. In other words, the persistence *gap*. Note, however, that a simple comparison of how women compare to men in terms of continuing with the application process after an initial rejection will fail to cleanly identify gender differences in persistence independent of gender differences in application characteristics. A comparison of gender-specific means may yield biased estimates because the grant of a patent is likely correlated with applicant characteristics, many of which are unlikely to be observed by the econometrician. The challenge for identifying both the differential patent grant rate for men and women is the problem of differential selection into various phases of patent examination process. If, for example, men file applications that are inherently more easily obtained, they may appear more persistent following a rejection. To estimate the extent to which patent outcomes depend on differential exit from the application process, we need variation in initial patent outcomes that are orthogonal to other determinants of patents, such as the merit of a given application.

We thus use an instrumental variables (IV) strategy that leverages exogenous variation in the likelihood of patent denial at the *first* stage of the patent prosecution that arises based on the random assignment of patent examiner. To obtain this type of variation, we use the random assignment of applications within the USPTO's review process to get as close as possible to the ideal experiment. Our design is similar in spirit to several recent studies about the patent prosecution process, such as Sampat and Williams (2019) and Farre-Mensa et al. (2017). The main intuition is that patent applications to the PTO are assigned to examiners quasi-randomly, and examiners differ systematically in their propensity to approve patents at any stage. Lenient examiners are more likely to grant a patent than are stricter examiners, holding the quality of the proposed invention constant. Thus, drawing a harsher examiner increase the probability of initial rejection, regardless of patent quality or other ex ante characteristics. We define examiner leniency as the leave-one-out initial rejection rate of examiner by unit-year (i.e., the proportion of all *other* applications for which a decision of initial rejection is made by a given examiner in each art unit - year):[2] [3]

$$Harshness_{ae} = \left(\frac{1}{n_e}\right)\left(\sum_{k \neq a}^{n_e} ER_k\right)$$

In this expression, e the examiner assigned to an application a, n_e is the total number of applications seen by examiner e in art unit- year, k indexes the applications seen by examiner e, and ER_k, an initial reject, is equal to one if the applicant did not receive a patent when the first response was given by examiner e for patent application k. We construct this measure for each application, so for a given application, the harshness measure captures how stringent an examiner is based on all other applications he or she reviews. This avoids any bias from the current application. Figure 2 shows that examiner rejection rates at the initial decision stage vary substantially within year and art unit.

Using this instrument, we can then cleanly estimate heterogeneity in patent prosecution persistence:

$$Y_a = \beta_1 \widehat{Initial\ Rej}_a + \beta_2 Female_a + \beta_3 [Female\ \widehat{\times\ Initial}\ Rej]_a + \mu_{ut} + \epsilon_a \tag{3.2}$$

where we instrument for $\widehat{Initial\ Rej}_a$ and $[Female\ \widehat{\times\ Initial}\ Rej]_a$ using $Harshness_e$ and $[Female_a\ \times\ Harshness_e]$, respectively.

Before proceeding, we provide visual evidence in favor of a strong first-stage relationship in Figure 3. We calculate the mean initial rejection rate for each examiner, residualized by Art Unit-by-application year fixed effects. When we relate this measure of examiner "leniency" to FOA outcomes, we observe a strong relationship. Contrast this strong correlation (close to a 45-degree relationship) to the other plot, where we regress the predicted probability of rejection controlling for several predetermined observable characteristics: the number of innovator applicants, the proportion of female innovators on an application, whether an application is assigned to an employer, etc. This relationship is displayed nonparametrically in the darker plot. There is no strong visual relationship between the predicted probability of rejection and our instrument. This figure thus provides indirect support of a strong first stage relationship, as well as of the exclusion restriction.

3.3 Results

3.3.1 Main Findings

Motivating Evidence

We begin our analysis by presenting suggestive evidence that persistence has a central role in explaining the gender innovation gap. Panel A of Table 3 presents the results of a simple OLS regression of the impact of gender on initial rejection rates, controlling for art

[2] We construct examiner harshness based on art unit- year because applications are assigned to examiners with art unit in a given year.

[3] In Appendix C.1, we discuss an alternative definition of examiner leniency, using the leave-one-out patent rejection rates. The main results are robust to using the alternative definition.

unit- year. The definition of Female changes across columns: Column (1) is the proportion of females on an application; Column (2) is includes only solo applicants and an indicator for female applicant; Column (3) is 50% or more females on an application; and Column (4) is 100% females on an application. We find suggestive evidence that female-led teams are marginally more likely to receive initial rejection compared to male-led teams. While statistically significant, the effects are economically small. For instance, in Column (2), which restricts attention to applications filed by individuals, we find that female inventors are 0.008 percentage points more likely to receive an initial rejection. We observe similar magnitudes across all specifications. Since innovator gender might be correlated with unobserved characteristics such as patent quality, we cannot role out that examiners are discriminating against female-led team. Nevertheless, the above evidence suggests that discrimination, at least at the FOA stage, is not a major driver of the innovation gender gap.

In contrast, observing Panel B in Table 3, we see that female is correlated with significant reductions in the probability of a patent being granted. For instance, in Column (2) individuals application filed by women are 0.051 percentage points less likely to receive a patent compared to individual applications filed by men. This magnitude is more than six time the impact of initial rejection. We observe similar magnitudes across all specifications. This suggests that there exists a gender gap in application accepted. More importantly, this gap to not driven by differences in initial rejection rates, but instead arises in subsequent stages of the patent application. This stylized fact is consistent with our hypothesis that heterogeneities in persistence drive the innovation gender gap.

Before turning to our main analysis, we also confirm that an adverse outcome in the initial stages, in general leads to the reduced likelihood of ultimately being awarded a patent. Panel A of Table 4 presents results based on the OLS regression in Equation 3.1, but estimated for men and women separately or pooled together. Both men and women (Columns 2-3) are less likely to complete the patent application process (successfully) if rejected at the initial stage – although men are less likely to exit the patent prosecution process than women: 28.5% of women drop out (i.e., fail to ever receive a patent), while only 25.5% of men drop out. However, as discussed above, these regressions potentially yield biased estimates of the likelihood of continuing the process. Persistence in the patent process at this point may be driven by higher quality applications. Moreover, the gender differences in attrition may be due to the fact that women are also less likely to submit poorer quality applications, which in turn lead examiners to reject them at higher rights. When we instead instrument for the likelihood of initial rejection with a given examiner's overall initial rejection rate, we find that results are similar in direction, and are in fact larger magnitude. Panel B confirms that an exogenous increase in the likelihood of initial rejection does in fact lead innovators to exit the patent process despite these the non-final nature of the rejection. Recall that no rejection is final, meaning that non-receipt of a patent right is the best measure of failing to persist in the patenting process. The above results indicate a substantial effect of receiving an initial rejection on subsequent incompletion of the application process.

Persistence by Gender We now turn to our primary question, which studies heterogeneity by gender in innovators' responses to initial rejections. This approach allows us to directly compare woman's to men's persistence in a unified regression framework. Columns (1)-(4) of Table 5 present our primary results for this paper. Each column indicates an estimation using the instrumental variable strategy in Equations 3.2. The definition of Female changes across columns and is indicated below each column, similar to Table 3.

Our primary outcome is whether an applicant/team proceeds to the next step of the application, i.e. files an amendment or appeal. Collectively, the results demonstrate that women and women-lead teams are significantly less likely to persist in the patent process if they receive an initial rejection compare to their men counterparts. This finding is consistent across a number of specifications. Column (1) summarizes the negative relationship between female-led teams and innovation across our entire sample: for every 10% increase in women on an application, the likelihood of dropping out prior to application re-submission increases by 3.6 percentage points (p.p.). Recall that because we are leveraging random variation in likelihood of rejection at this stage, these estimates avoid potential bias from unobservable application characteristics.

In Column (2), we limit our ample to applications submitted by individuals. This approach, while reducing our sample size, provides arguably the cleanest estimate of gender persistence disparities in this innovation setting. Examining individual inventors allows us to avoid issues related to selection in team composition. The Column (2) estimates are similar in magnitude, and suggest that a woman who applies for a patent is about 4 p.p. less likelihood than men applying within the same art unit to amend, resubmit, or appeal if her initial set of patentability claims is rejected at the initial stage. Our estimate suggests that an exogenous rejection reduces the percentage of women applicants after the initial examination by 5% (= 0.043 / 0.86).

Columns (3) and (4) provide estimates of the primary specification using different definitions of "female-led patents" that facilitate ease of interpretation. The estimates again suggest that women are less likely than men to persist beyond an initial rejection in the patent examination process. In column (3), we observe that when patent teams that are majority female (patents in which 50% or more of the inventors on an application are women) receive an initial rejection, these teams are 3 p.p. less likely to continue with the application process than applicant teams that are majority male (a reduction of 4% off the mean). Finally, all-female teams are at least 7 p.p. less likely to continue with a patent application after an initial rejection. The apparent monotonic relationship between fraction woman and patent application persistence collectively provides compelling evidence that women are deferentially deterred from continuing in the patent process after an initial rejection. This finding is consistent with ample evidence from other high-stakes labor markets, such as in politics (Wasserman, 2018).

Impact of Patent Approval We now turn to study the downstream effects of persistence on receive patents. The results just discussed suggest the differential effects of initial rejection on immediate drop-out. However, it is unclear whether this mechanism explains the overall gender disparities in innovation. We thus also examine whether women are differentially deterred from ultimately *completing* patent applications after (exogenously-determined) initial rejections by their assigned patent examiners. Successful patent grants may involve several examiner rejections of specific claims, followed by applicant amendments, before a patent is finally awarded.

The results are presented in Table 6. For brevity, we focus our discussion just on applications in which over 50% of the team is comprised of women, as presented in Column (3). These teams are 5.5 p.p. less likely to receive patents than patent application teams that are majority men. Given that these teams were 3 p.p. more likely to drop out immediately

(Column (3) in Table 5)– i.e., without refiling an amended application – our results suggest that about 60% of the overall gender patent granting gap is explained by women's differential deterrence when an examiner makes her initial determination. Reassuringly, we observe similar magnitude differences between immediate and downstream gender differences across our measures of gender application composition. In all cases, the differential deterrence of women after initial rejection is larger for final patent completion than for completing the immediate next step of the process, and the ratio of is centered around 60%.

3.3.2 Sources of Heterogeneity

We turn now to sources of heterogeneity in women's relative lack of persistence in the patent prosecution process relative to men. First, we consider whether working for firms (rather than independently) dampens women's greater tendency to exit the patent process after an initial rejection. To test this, we use the "assignments" of applications from employee inventors to the firms that employ them. A firm "assignment" indicates that a firm will assume the ownership right if granted, and is thus an indication of firm backing an inventor team.[4] In addition to increasing the overall likelihood that an innovator team will proceed beyond the initial phase of the patent examination reasons, there are strong conceptual reasons to hypothesize that firm backing may ameliorate some of the persistence gap in our setting. In particular, firms are more likely to supply mentorship to women, which has been shown to increase persistence in STEM careers (Blau et al., 2010).

Table 7 demonstrates that working within firms does offset women differential deterrence after an initial rejection. Starting with Columns (1)-(2), we can first observe that firm-backed patent applications are considerably more likely to proceed beyond an initial rejection for both men and women.[5] Thus, that relative to women teams applying independently female innovator teams from

firms are more likely to continue the patent application process in some manner after a receiving a (quasi-exogenous) rejection. Moreover, examination of the coefficient on Female × Initial Rejection × Firm suggests that the added persistence benefit for women applying as part of a firm exceeds that of men applying in firms by 3.8 p.p. (5.2 p.p.) for half-female (full female) innovator teams. However, majority-women application teams backed by firms are still less likely to either appeal or amend their applications after an rejections compared to majority-men teams backed by firms; This result required the summation of three coefficients, and is displayed below each column. Thus, firm backing narrows the persistence gap between women and men-led teams, but does not erase it completely. Similar results are born out in Columns (3)-(4) in terms of persistence through the final patent issued stage.

Finally, we examine the impact of examiner's gender on heterogeneity in gender persistence. The results are presented in Table 8. We find little clear evidence that examiner's gender has a bearing on the persistence gap between men and women.

---

[4]Legally, the "original applicant is presumed to be the owner of an application for an original patent, and any patent that may issue therefrom, unless there is an assignment" (Graham et al., 2015).

[5]The coefficient Initial Rejection × Firm captures the effect for men, and Initial Rejection × Firm + Female × Firm + Female × Initial Rejection × Firm (not presented in table) captures the effect for women. Both are positive across al specifications.

3.4 Conclusion

In this study we seek to identify the extent to which gender differences in deterrence after early setbacks may contribute to gender differences in STEM performance. This work make an important contribution to the literature on persistence and innovation by opening the black box of gender disparities in innovation and identifying a key reason for this gap. We identify that persistence within the patent process, primarily in its initial stages, drives differential outcomes in patenting. This gap widens as the presence of women on inventor teams increases. We identify potential interventions that policymakers can consider to begin addressing the gender gap in patenting, so that future innovations better serve the needs of a diverse and varied world.

Figures

Figure 1: Patent Process

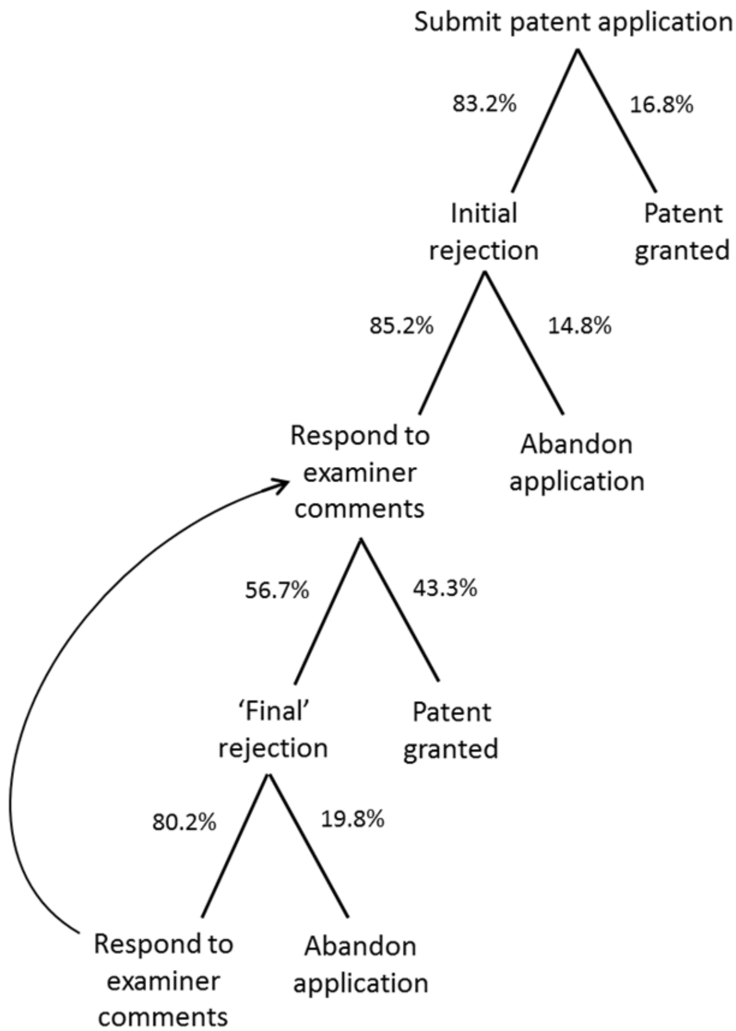
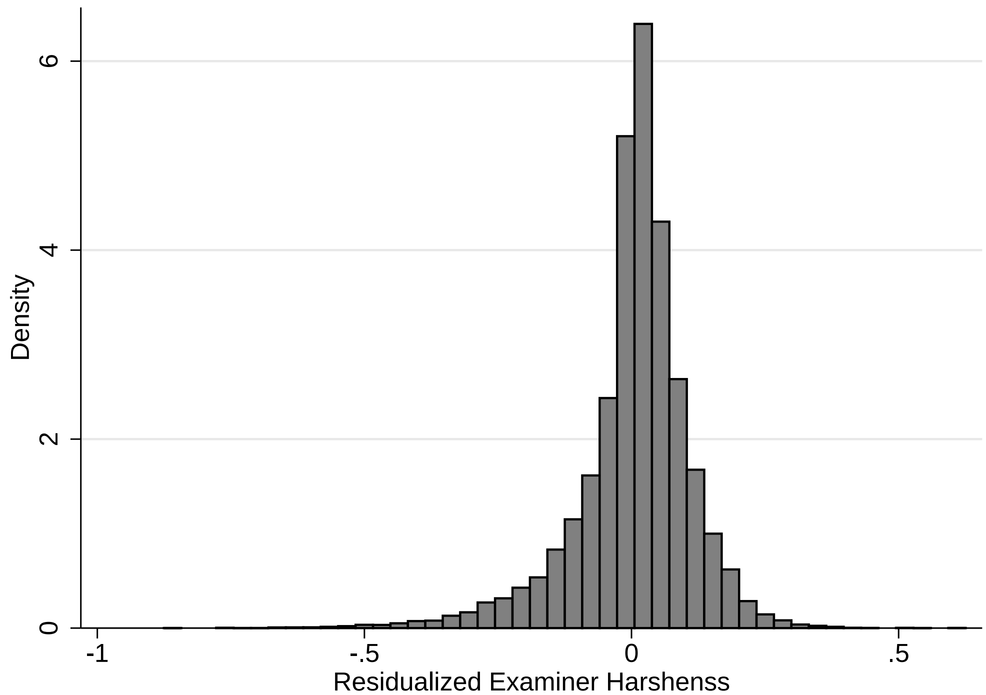
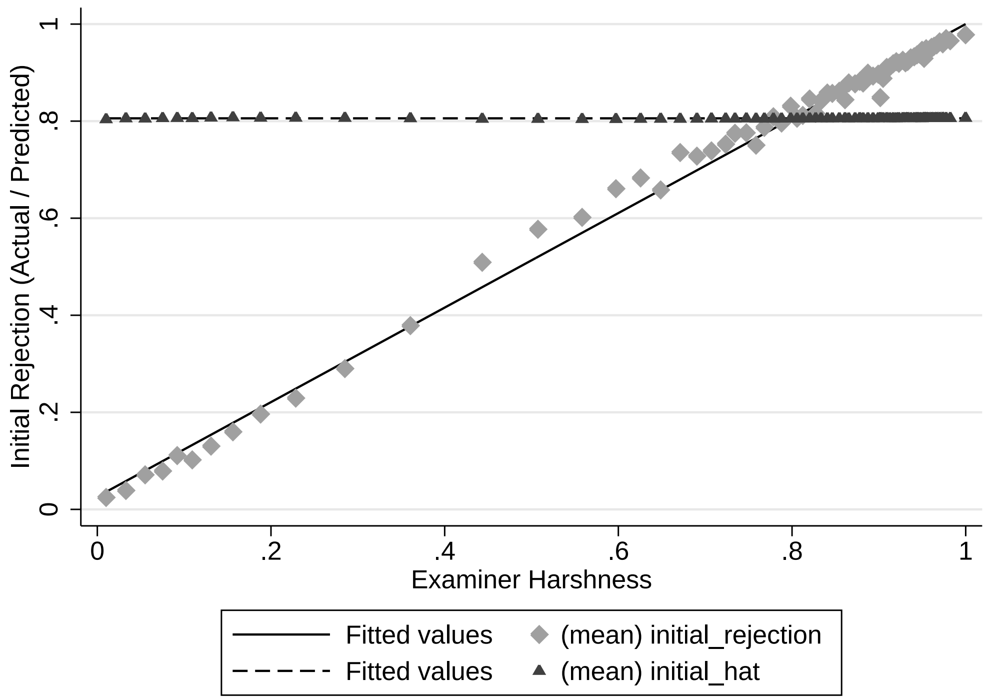


Figure 2: Distribution of Examiner Harshness by Initial Rejection (Residualized)



Note: This figure shows the distribution of patent initial rejection rates, residualizing by the full set of art-unit-by-application-year fixed effects.

Figure 3: Examiner Harshness and (Predicted) Initial Rejection



Note: This figure relates our examiner leniency measure by Art Unit-by-application year to two variables: (1) the initial rejection rate and (2) the *predicted* rejection rate, where we predict a rejection as a function of predetermined observable characteristics: the number of innovator applicants, the proportion of female innovators on an application, whether an application is assigned to an employer, etc.

Tables

Table 1: Summary Statistics

	Mean	SD	Min	Max
Applications (N=1,350,345)				
Small Entity Indicator	0.39	0.49	0	1
Employer Assignment	0.62	0.48	0	1
patent_issued	0.67	0.47	0	1
Number of Team Members	2.12	1.42	1	10
Solo Inventors	0.45	0.50	0	1
Proportion of Female Team Members	0.086	0.23	0	1
>=1 woman on team	0.16	0.37	0	1
>=50% women on team	0.091	0.29	0	1
All-femaleteam	0.041	0.20	0	1
Individual female inventor	0.036	0.19	0	1
Number of Initial Rejections	1.23	0.97	0	14
Number of Initial Appeals	1.24	1.25	0	24
Number of Final Rejections	0.58	0.84	0	12
Number of Final Appeals	0.62	1.09	0	24

Note: Words Words Words

Table 2: Probability of Submitting Amendments and Outcomes

	Submitted Initial Amendment	Submitted Final Amendment	Patent Issued
Full sample	0.86	0.80	0.67
Solo Male	0.83	0.77	0.66
Solo Female	0.78	0.71	0.61
Teams with >= 1 woman	0.85	0.78	0.62
Teams with >=50% women	0.82	0.75	0.61
All-femaleteams	0.78	0.72	0.61

Note: Initial and final amendment probabilities are calculated for the sample of applications that received initial and final rejections, respectively.

Table 3: Motivating Evidence - Effect of Gender on Patent Application Outcomes

	(1)	(2)	(3)	(4)
	Panel A: Effect of Gender on Initial Rejection			
Female	0.007***	0.008***	0.004***	0.006***
	(0.001)	(0.002)	(0.001)	(0.002)
Dependent Var. Mean	0.80	0.77	0.80	0.80
	Panel B: Effect of Gender on Patent Granted			
Female	-0.052***	-0.051***	-0.042***	-0.070***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1031848	478987	1031848	1031848
# of Clusters	38682	38522	38682	38682
Dependent Var. Mean	0.70	0.69	0.70	0.70
Female Definition	Proportion	Solo	Half Female	All Female

Note: This table reports regression coefficients from eight separate regressions. An observation is a patent application. Outcomes are denoted at sub-headers. Definitions of the independent variable are denoted below each column and are described in the text. All regressions include art unit-year fixed effect. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Effect of Initial Rejection on Patent Issuance- Baseline Specification

	All Applicants	Female Only	Male Only
	(1)	(2)	(3)
	Panel A: OLS Regressions		
Initial Rejection	-0.259***	-0.284***	-0.255***
	(0.002)	(0.007)	(0.002)
	Panel B: Instrumental Variable Regressions		
Initial Rejection	-0.677***	-0.757***	-0.668***
	(0.011)	(0.038)	(0.011)
Observations	1031848	45184	869393
# of Clusters	38682	16318	38682
Dependent Var. Mean	0.70	0.64	0.71

Note: This table reports regression coefficients from six separate regressions. An observation is a patent application. Sample included in each regression is indicated in the columns' titles. Outcomes is a dummy variable for whether a patent was issued. Panel A presents OLS regressions. Panel B instruments for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. All regressions include art unit-year fixed effect. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Effect of Initial Rejection on Next Step

	(1)	(2)	(3)	(4)
Female X Initial Rejection	-0.036***	-0.043***	-0.031***	-0.072***
	(0.003)	(0.004)	(0.003)	(0.004)
Observations	1031848	478987	1031848	1031848
# of Clusters	38682	38522	38682	38682
Dependent Var. Mean	0.88	0.86	0.88	0.88
Female Definition	Proportion	Solo	Half Female	All Female

Note: This table reports regression coefficients from six separate regressions. An observation is a patent application. Outcomes is a dummy variable for whether application was continued after initial rejection. Columns (5)-(6) use an alternative definition of the outcome variable, as described in the text. We instrument for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Effect of Initial Rejection on Patent Granted

	Main Definition			
	(1)	(2)	(3)	(4)
Female X Initial Rejection	-0.068***	-0.070***	-0.055***	-0.100***
	(0.005)	(0.006)	(0.004)	(0.006)
Observations	1031848	478987	1031848	1031848
# of Clusters	38682	38522	38682	38682
Dependent Var. Mean	0.70	0.69	0.70	0.70
Female Definition	Proportion	Solo	Half Female	All Female

Note: This table reports regression coefficients from four separate regressions. An observation is a patent application. Outcomes is a dummy variable for whether a patent was issued. We instrument for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Heterogeneity by Employment Status

	Next Step		Patent Issued	
	(1)	(2)	(3)	(4)
Female × Initial Rejection × Firm	0.038***	0.052***	0.038***	0.054***
	(0.005)	(0.007)	(0.008)	(0.011)
Initial Rejection × Firm	0.126***	0.126***	0.147***	0.147***
	(0.002)	(0.002)	(0.004)	(0.004)
Female × Initial Rejection	-0.042***	-0.070***	-0.064***	-0.095***
	(0.004)	(0.005)	(0.007)	(0.008)
Female × Firm	-0.008***	-0.011***	-0.010	-0.019**
	(0.003)	(0.003)	(0.006)	(0.008)
Observations	1031848	1031848	1031848	1031848
# of Clusters	38682	38682	38682	38682
Dependent Var. Mean	0.88	0.88	0.70	0.70
$\beta 1 + \beta 3 + \beta 4$	-0.012	-0.029	-0.036	-0.059
Pvalue	0.000	0.000	0.000	0.000
Female Definition	Half Female	All Female	Half Female	All Female

Note: This table reports regression coefficients from four separate regressions. An observation is a patent application. Outcomes are denoted in column titles and described in the text. We instrument for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. Firm is a dummy variable for whether the patent is assigned to a firm. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Heterogeneity by Examiner Gender

	Next Step		Patent Issued	
	(1)	(2)	(3)	(4)
Female × Initial Rejection × Examiner Female	-0.004	-0.002	-0.000	0.012
	(0.006)	(0.008)	(0.010)	(0.013)
Initial Rejection × Examiner Female	-0.003	-0.003	-0.035***	-0.036***
	(0.003)	(0.002)	(0.006)	(0.006)
Female × Initial Rejection	-0.032***	-0.075***	-0.056***	-0.106***
	(0.004)	(0.006)	(0.007)	(0.009)
Female × Examiner Female	0.003	-0.001	0.003	-0.003
	(0.003)	(0.004)	(0.007)	(0.009)
Observations	865815	865815	865815	865815
# of Clusters	31766	31766	31766	31766
Dependent Var. Mean	0.87	0.87	0.70	0.70
Female Definition	Half Female	All Female	Half Female	All Female

Note: This table reports regression coefficients from four separate regressions. An observation is a patent application. Outcomes are denoted in column titles and described in the text. We instrument for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. Examiner Female is a dummy variable for whether patent examiner is a female. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Bibliography

Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl (2009). The effects of entry on incumbent innovation and productivity. *The Review of Economics and Statistics 91*(1), 20–32.

Alan, S., S. Ertac, E. Kubilay, and G. Loranth (2020). Understanding gender differences in leadership. *The Economic Journal 130*(626), 263–289.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016). University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review 106*(3), 525–62.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics 37*(3), 668–691.

Armstrong, M., J. Vickers, and J. Zhou (2009). Prominence and consumer search. *The RAND Journal of Economics 40*(2), 209–233.

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. *Handbook of Economic Field Experiments 1*, 73–140.

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

Bar-Isaac, H., S. Tadelis, et al. (2008). Seller reputation. *Foundations and Trends® in Microeconomics 4*(4), 273–351.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management 17*(1), 99–120.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management science 15*(5), 215–227.

Beede, D. N., T. A. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms (2011). Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief* (04-11).

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica*, 841–890.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, 889–917.

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

Blau, F. D., J. M. Currie, R. T. Croson, and D. K. Ginther (2010). Can mentoring help female assistant professors? interim results from a randomized trial. *American Economic Review 100*(2), 348–52.

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

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

Brands, R. A. and I. Fernandez-Mateo (2017). Leaning out: How negative recruitment experiences shape women's decisions to compete for executive roles. *Administrative Science Quarterly 62*(3), 405–442.

Bresnahan, T. F. and P. C. Reiss (1991). Entry and competition in concentrated markets. *Journal of political economy 99*(5), 977–1009.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, 955–969.

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

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

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

Caillaud, B. and B. Jullien (2003). Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics*, 309–328.

Cao, G., G. Z. Jin, X. Weng, and L.-A. Zhou (2018). Market expanding or market stealing? competition with network effects in bike-sharing. *Working paper , National Bureau of Economic Research*.

Cennamo, C., Y. Gu, and F. Zhou (2016). Value co-creation and capture in the creative industry: The u.s. home video game market. *Working paper*.

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

Corts, K. S. and M. Lederman (2009). Software exclusivity and the scope of indirect network effects in the us home video game market. *international Journal of industrial Organization 27*(2), 121–136.

Cullen, Z. and C. Farronato (2019). Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Working paper*.

Cusumano, M. A., A. Gawer, and D. B. Yoffie (2019). *The business of platfroms*. HarperCollins-Business London.

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

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

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science 49*(10), 1407–1424.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American economic review 67*(3), 297–308.

Dubé, J.-P. H., G. J. Hitsch, and P. K. Chintagunta (2010). Tipping and concentration in markets with indirect network effects. *Marketing Science 29*(2), 216–249.

Ellison, G. and S. F. Ellison (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics 3*(1), 1–36.

Farre-Mensa, J., D. Hegde, and A. Ljungqvist (2017). What is a patent worth? evidence from the us patent “lottery”. *The Journal of Finance*.

Farrell, J. and G. Saloner (1985). Standardization, compatibility and innovation. *The RAND Journal of Economics 16*, 70–83.

Farronato, C. and A. Fradkin (2018). The welfare effects of peer entry in the accommodation market: The case of airbnb. Technical report, National Bureau of Economic Research.

Fosfuri, A. and M. S. Giarratana (2009). Masters of war: Rivals’ product innovation and new advertising in mature product markets. *Management Science 55*(2), 181–191.

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

Fradkin, A., E. Grewal, and D. Holtz (2019). Reciprocity in two-sided reputation systems: Evidence from an experiment on airbnb.

Fudenberg, D. and J. Tirole (1984). The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *American Economic Review 74*(2), 361–366.

Gawer, A. and R. Henderson (2007). Platform owner entry and innovation in complementary markets: Evidence from intel. *Journal of Economics & Management Strategy 16*(1), 1–34.

Goolsbee, A. and C. Syverson (2008). How do incumbents respond to the threat of entry? evidence from the major airlines. *The Quarterly Journal of Economics 123*(4), 1611–1633.

Graham, S. J., A. C. Marco, and R. Miller (2015). The uspto patent examination research dataset: A window on the process of patent examination. *Georgia Tech Scheller College of Business Research Paper No. WP 43*.

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review 29*(6), 911–922.

Hagiu, A. and J. Wright (2015). Multi-sided platforms. *International Journal of Industrial Organization 43*, 162–174.

Hauser, J. R. and S. M. Shugan (2008). Defensive marketing strategies. *Marketing Science 27*(1), 88–110.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

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

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

Hollenbeck, B., S. Moorthy, and D. Proserpio (2019). Advertising strategy in the presence of reviews: An empirical analysis. *Forthcoming Marketing Science*.

Hortaçsu, A. and C. Syverson (2007). Cementing relationships: Vertical integration, foreclosure, productivity, and prices. *Journal of Political Economy 115*(2), 250–301.

Horton, J. and J. Golden (2018). Reputation inflation: Evidence from an online labor market. *New York University, Working Paper 1*.

Hotelling, H. (1929). Stability in competition. *The Economic Journal 39*(153), 41–57.

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.

Illanes, G. and S. Moshary (2018). Market structure and competition: Evidence from a natural experiment in liquor licensure. *Working paper*.

Jensen, K., B. Kovács, and O. Sorenson (2018). Gender differences in obtaining and maintaining patent rights. *Nature biotechnology 36*(4), 307–309.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

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

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

Jin, G. Z. and M. Rysman (2015). Platform pricing at sports card conventions. *The Journal of Industrial Economics 63*(4), 704–735.

Kadiyali, V. (1996). Entry, its deterrence, and its accommodation: A study of the us photographic film industry. *The Rand Journal of Economics*, 452–478.

Kaiser, U. and J. Wright (2006). Price structure in two-sided markets: Evidence from the magazine industry. *International Journal of Industrial Organization 24*(1), 1–28.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *American economic review 75*(3), 424–440.

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

Koning, R., S. Samila, and J.-P. Ferguson (2019). Female inventors and inventions. *Available at SSRN 3401889*.

Kovács, B. and A. J. Sharkey (2014). The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly 59*(1), 1–33.

Kuppuswamy, V. and E. R. Mollick (2016). Second thoughts about second acts: Gender differences in serial founding rates. *Available at SSRN 2752689*.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review 103*(7), 2960–3000.

Lemley, M. A. and B. Sampat (2008). Is the patent office a rubber stamp. *Emory LJ 58*, 181.

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

Li, Z. and A. Agarwal (2016). Platform integration and demand spillovers in complementary markets: Evidence from facebook's integration of instagram. *Management Science 63*(10), 3438–3458.

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

Mahajan, V., S. Sharma, and R. D. Buzzell (1993). Assessing the impact of competitive entry on market expansion and incumbent sales. *Journal of marketing 57*(3), 39–52.

Marshall, A. (1890). Principles of economics.

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

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica 71*(6), 1695–1725.

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

Milli, J., E. Williams-Baron, M. Berlan, J. Xia, and B. Gault (2016). Equity in innovation: women inventors and patents. *Institute for Women's Policy Research Report*.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989). Industrialization and the big push. *Journal of political economy 97*(5), 1003–1026.

Nair, H., P. Chintagunta, and J.-P. Dubé (2004). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics 2*(1), 23–58.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, 395–421.

Nickell, S. J. (1996). Competition and corporate performance. *Journal of Political Economy 104*(4), 724–746.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics 122*(3), 1067–1101.

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

Ohashi, H. (2003). The role of network effects in the us vcr market, 1978–1986. *Journal of Economics & Management Strategy 12*(4), 447–494.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from chilean plants. *The Review of Economic Studies 69*(1), 245–276.

Porter, M. E. (1989). How competitive forces shape strategy. In *Readings in strategic management*, pp. 133–143. Springer.

Porter, M. E. (1997). Competitive strategy. *Measuring business excellence 1*(2), 12–17.

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

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

Rochet, J.-C. and J. Tirole (2006). Two-sided markets: a progress report. *The RAND journal of economics 37*(3), 645–667.

Rysman, M. (2004). Competition between networks: A study of the market for yellow pages. *The Review of Economic Studies 71*(2), 483–512.

Sampat, B. and H. L. Williams (2019). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

Samuelson, P. A. (1951). *Economics-an introductory analysis.* Mcgraw-Hill Book Company, Inc.; New York; London; Toronto.

Sarada, S., M. J. Andrews, and N. L. Ziebarth (2019). Changes in the demographics of american inventors, 1870–1940. *Explorations in Economic History 74*, 101275.

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

Seamans, R. and F. Zhu (2013). Responses to entry in multi-sided markets: The impact of craigslist on local newspapers. *Management Science 60*(2), 476–493.

Seamans, R. C. (2013). Threat of entry, asymmetric information, and pricing. *Strategic Management Journal 34*(4), 426–444.

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

Shen, Q. and P. Xiao (2014). Mcdonald's and kfc in china: Competitors or companions? *Marketing Science 33*(2), 287–307.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. *Econometrica 49*(1), 105.

Springel, K. (2018). Network externality and subsidy structure in two-sided markets: Evidence from electric vehicle incentives. *Working Paper*.

Stewart Stute, S. (2019). The gender gap in patents: An exploration of bias against women in patent attainment and "blockchain" as potential remedy. *Working Paper*.

Stigler, G. J. (1961). The economics of information. *Journal of political economy 69*(3), 213–225.

Stremersch, S., G. J. Tellis, P. Hans Franses, and J. L. Binken (2007). Indirect network effects in new product growth. *Journal of Marketing 71*(3), 52–74.

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222.

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

Tversky, A. and D. Kahneman (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Wasserman, M. (2018). Gender differences in politician persistence. *Available at SSRN 3370587*.

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

Young, A. (2016). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *London School of Economics, Working Paper*.

Zbaracki, M. J., M. Ritson, D. Levy, S. Dutta, and M. Bergen (2004). Managerial and customer costs of price adjustment: direct evidence from industrial markets. *Review of Economics and statistics 86*(2), 514–533.

Zhu, F. and M. Iansiti (2012). Entry into platform-based markets. *Strategic Management Journal 33*(1), 88–106.

Zhu, F. and Q. Liu (2018). Competing with complementors: An empirical look at amazon. com. *Strategic Management Journal 39*(10), 2618–2642.

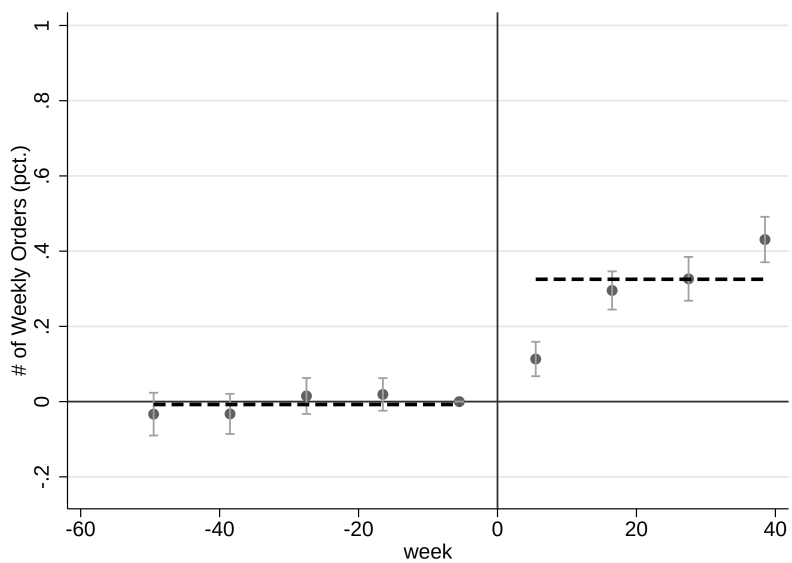
Appendices

Appendix A

Smaller Slices of a Growing Pie: The Effects of Entry in Platform Markets

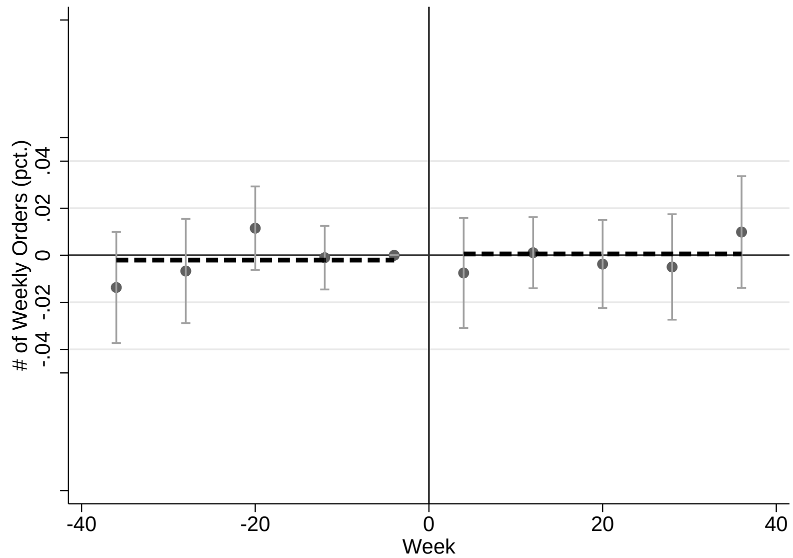
A.1 Appendix Figures and Tables

Figure A1: Percentage Change in Market-Level Number of Orders

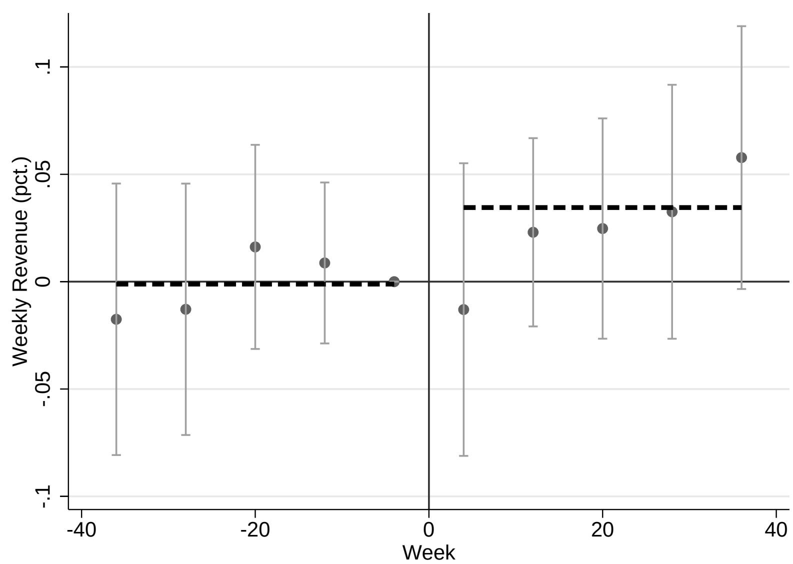


Note: This figure presents event-time estimates from a version of equation 1.4. The dependent variables are the inverse hyperbolic sine transformation, and should be interpreted as percentage changes. The unit of observation is city-week, including both incumbent and newly added businesses. The dots represent point estimates from regressing the dependent variable on a treatment indicator interacted with 9-weeks bins, and city and week-state fixed effects. The treatment indicator compares cities that experienced almost no change in the percentage of businesses available on the platform to cities that experienced meaningful changes. The coefficient in the first period prior to implementation is normalized to zero. The vertical bar represent 95% confidence intervals, where standard errors are clustered at the city level.

Figure A2: Impact of Entry on Incumbent Firms



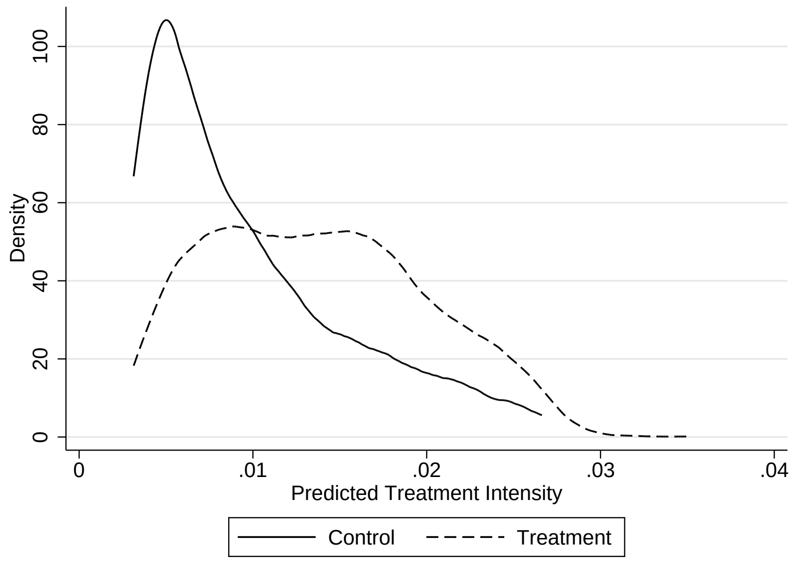
(a) Change in Weekly Number of Orders (Percentage)



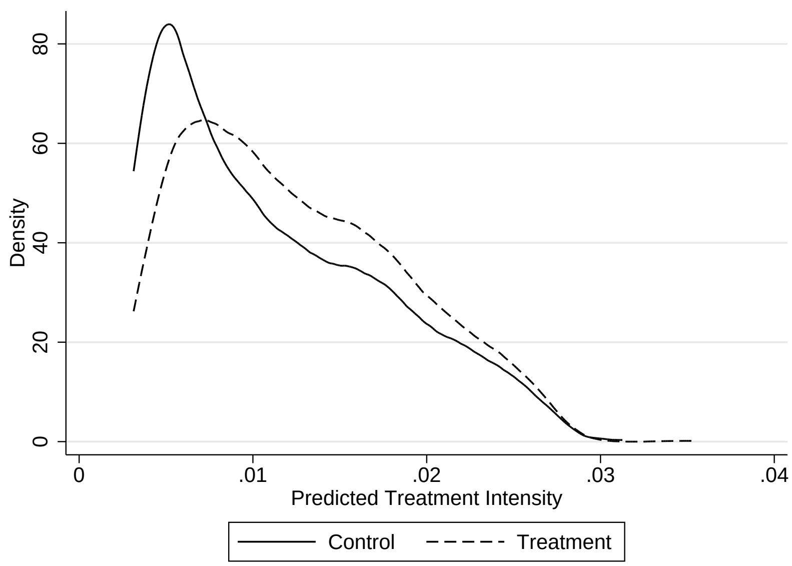
(b) Change in Weekly Revenue (Percentage)

Note: This figure presents event-time estimates from a version of equation 1.4. The dependent variables are the inverse hyperbolic sine transformations of weekly number of orders (Panel A) and weekly revenue (Panel B), and should be interpreted as percentage changes. The unit of observation is city-week. The dots represent point estimates from regressing the dependent variable on a treatment indicator interacted with 9-weeks bins, and city and week-state fixed effects. The treatment indicator compares cities that experienced almost no change in the percentage of businesses available on the platform to cities that experienced meaningful changes. The coefficient in the first period prior to implementation is normalized to zero. The vertical bar represent 95% confidence intervals, where standard errors are clustered at the city level.

Figure A3: Predicted Treatment Intensity by Treatment Assignment



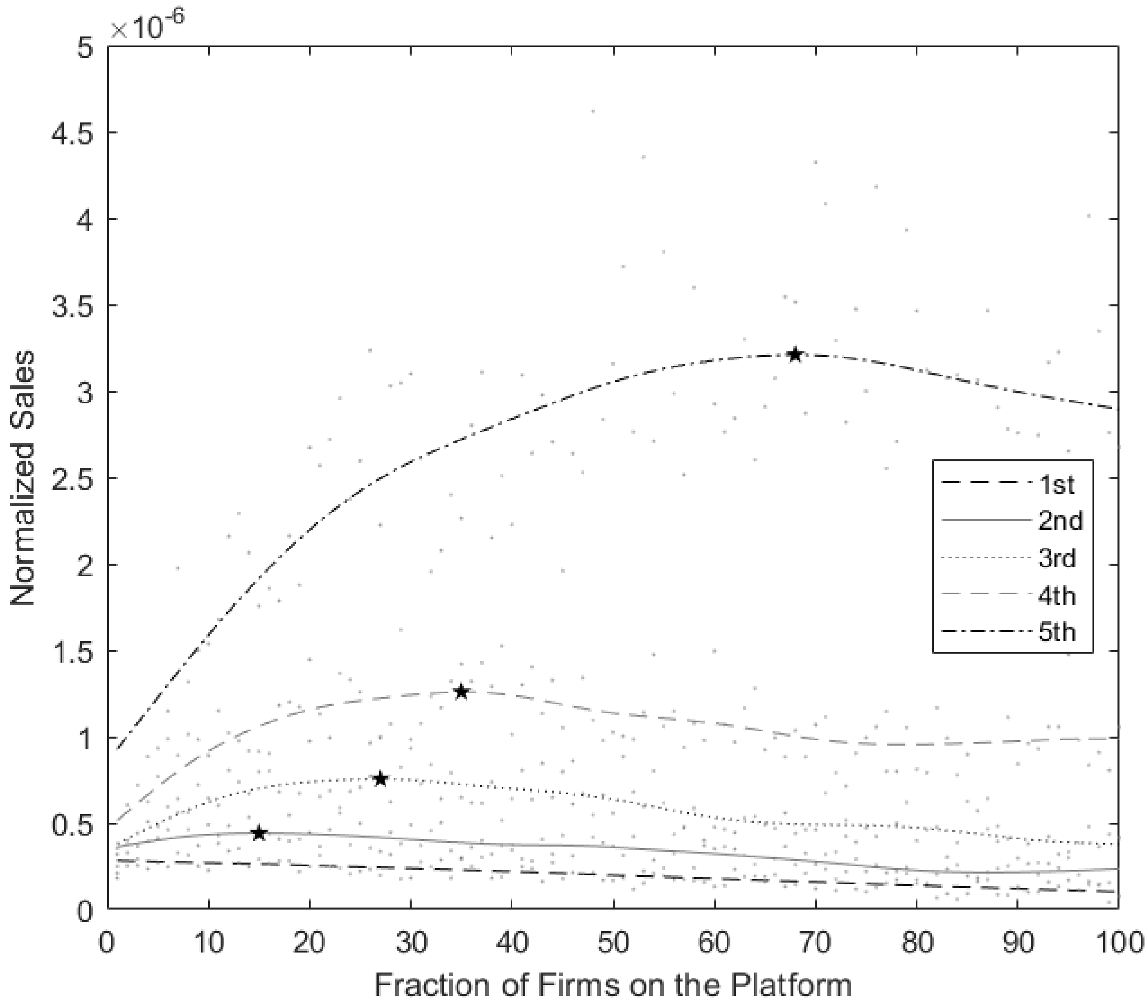
(a) Original Distributions



(b) Distribution Using Within Bin Assignment

Note: The figure presents the distribution of propensity score by treatment intensity. Propensity score are estimated on the continuous change in share of restaurants on YTP. Treatment is an indicator for above median treatment intensity. Panel A presents the distributions of propensity scores in the original data by treatment assignment. Panel B presents the distributions of propensity scores when treatment status is assigned by propensity score bins.

Figure A4: Simulations of Firms' Sales by Rating Quantile- Including Demographics



Note: This figure present the change in number of sales by rating quantile as the share of firms on the platform grows. The parameters used to simulate the data are presented in column 4 of Table 12. The simulation algorithm is described in appendix A.5. Each gray dot represent the average over one thousand simulations, and the dashed and solid line are the smoothing spline by rating quantile. The stars mark the maximum of each smoothing spline.

Table A1: Placebo Trends

	(1)	(2)	(3)
	Panel A: Weekly Orders		
Treat*Post	-0.008	-0.003	-0.003
	(0.007)	(0.009)	(0.013)
Treat*Post*Low		-0.010	0.007
		(0.012)	(0.015)
$\beta1 + \beta2$		-0.013	0.004
Pvalue		0.179	0.686
	Panel B: Weekly Revenue		
Treat*Post	-0.020	-0.017	-0.022
	(0.018)	(0.023)	(0.033)
Treat*Post*Low		-0.010	0.047
		(0.031)	(0.040)
Observations	1477208	1477208	740706
# of Clusters	2729	2729	2625
$\beta1 + \beta2$		-0.027	0.025
Pvalue		0.301	0.363
Treatment Def.	25<>75	25<>75	25<>75
Quality Def.		Median	25<>75

Note: This table reports regression coefficients from 9 separate regressions, 3 per panel. An observation is business-week. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders (Panel A) and weekly-revenue (Panel B), and should be interpreted as percentage changes. Post is counterfactually set to the middle of the pre-treatment period. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A2: Placebo Outcomes

	(1)	(2)	(3)	(4)	(5)
	Panel A: Weekly Number of New Business				
Treat*Post	0.003	0.001	0.077		
	(0.003)	(0.004)	(0.084)		
Observations	321980	221935	321980		
# of Clusters	3788	2611	3788		
	Panel B: Weekly Number of Review				
Treat*Post	-0.001	-0.001	0.025	0.000	-0.000
	(0.001)	(0.001)	(0.034)	(0.001)	(0.002)
Treat*Post*Low				-0.002	-0.000
				(0.003)	(0.003)
Observations	31833645	17587953	31833645	17587953	10751909
# of Clusters	3964	2781	3964	2781	2047
$\beta1 + \beta2$				-0.002	-0.001
Pvalue				0.365	0.689
	Panel C: Average Weekly Rating				
Treat*Post	-0.000	-0.002	0.021	-0.001	-0.000
	(0.001)	(0.001)	(0.028)	(0.002)	(0.002)
Treat*Post*Low				-0.003	-0.004
				(0.002)	(0.003)
Observations	10236710	5685666	10236710	5685666	3444919
# of Clusters	3964	2781	3964	2781	2047
$\beta1 + \beta2$				-0.003	-0.005
Pvalue				0.049	0.009
Treatment Def.	Median	25<>75	Change	25<>75	25<>75
Quality Def.				Median	25<>75

Note: This table reports regression coefficients from 13 separate regressions, 3 in Panel A and 5 in Panels B and C. An observation is business-city in Panel A, and business-week in Panels B and C. The sample includes only non-YTP affiliated businesses and users. The dependent variables are inverse hyperbolic sine transformations and should be interpreted as percentage changes. Outcomes are indicated in the sub-headers and described further in the text. The sum of the coefficient is presented below each panel along with the corresponding p-value. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A3: Sensitivity to Geographical Market Definition

	By 5-Digits Zip Code				By County			
	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue	(5) Orders	(6) Orders	(7) Revenue	(8) Revenue
Treat*Post	0.049***	0.066***	0.136***	0.180***	0.048*	0.049**	0.094*	0.087
	(0.008)	(0.011)	(0.019)	(0.028)	(0.019)	(0.018)	(0.045)	(0.047)
Treat*Post*Low	-0.076***	-0.108***	-0.158***	-0.234***	-0.100***	-0.111***	-0.184**	-0.202***
	(0.011)	(0.014)	(0.026)	(0.034)	(0.023)	(0.022)	(0.057)	(0.058)
Observations	2268248	1283178	2268248	1283178	1613929	1273658	1613929	1273658
# of Clusters	6666	4197	6666	4197	1260	921	1260	921
$\beta 1 + \beta 2$	-0.027	-0.042	-0.022	-0.054	-0.052	-0.062	-0.090	-0.114
Pvalue	0.000	0.000	0.232	0.024	0.000	0.000	0.015	0.004
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. Geographic market definition indicated in sub-headings and described further in the text. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A4: Sensitivity to Alternative Market and Treatment Intensity Definitions

	By Market X Food Category				Alternative Treatment Definition			
	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue	(5) Orders	(6) Orders	(7) Revenue	(8) Revenue
Treat*Post	0.030**	0.027*	0.067**	0.072*	0.047***	0.054***	0.115***	0.137***
	(0.010)	(0.012)	(0.024)	(0.028)	(0.010)	(0.013)	(0.022)	(0.033)
Treat*Post*Low	-0.073***	-0.075***	-0.137***	-0.161***	-0.069***	-0.100***	-0.149***	-0.221***
	(0.012)	(0.015)	(0.030)	(0.035)	(0.018)	(0.016)	(0.036)	(0.043)
Observations	1804041	1084294	1804041	1084294	2173127	781414	2173127	781414
# of Clusters	8540	6708	8540	6708	3863	2686	3863	2686
$\beta 1 + \beta 2$	-0.043	-0.047	-0.070	-0.089	-0.021	-0.046	-0.034	-0.085
Pvalue	0.000	0.000	0.000	0.000	0.071	0.000	0.153	0.003
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. Market and treatment intensity definitions indicated in sub-headings and described further in the text. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A5: Sensitivity to Outliers

	Dropping Top & Bottom 5% of Cities				Randomization Inference			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Orders	Orders	Revenue	Revenue	Orders	Orders	Revenue	Revenue
Treat*Post	0.032^{**}	0.033^{*}	0.082^{**}	0.090^{*}	0.050^{***}	0.065^{***}	0.119^{***}	0.158^{***}
	(0.011)	(0.016)	(0.028)	(0.040)	[0.000]	[0.000]	[0.000]	[0.000]
Treat*Post*Low	-0.058^{***}	-0.060^{**}	-0.131^{***}	-0.134^{*}	-0.098^{***}	-0.119^{***}	-0.201^{***}	-0.250^{***}
	(0.015)	(0.021)	(0.039)	(0.056)	[0.000]	[0.000]	[0.000]	[0.000]
Observations	832690	423002	832690	423002	2173124	1321540	2173124	1321540
# of Clusters	2483	1422	2483	1422	3862	2714	3862	2714
$\beta1 + \beta2$	-0.026	-0.027	-0.049	-0.044				
Pvalue	0.016	0.082	0.089	0.302				
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. Columns 1 through 4 exclude the outliers cities. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. In columns 1 through 4 standard errors are in parentheses and are clustered at the city level. In columns 4 through 8 randomization inference p-values based on 2000 draws are reported in square brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A6: Sensitivity to Initial Differences

	Propensity Score Weighting				Blocked Treatment Assignment			
	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue	(5) Orders	(6) Orders	(7) Revenue	(8) Revenue
Treat*Post	0.044**	0.066***	0.105**	0.152***	0.033***	0.063***	0.090***	0.167***
	(0.014)	(0.016)	(0.034)	(0.039)	(0.009)	(0.014)	(0.021)	(0.033)
Treat*Post*Low	-0.067***	-0.089***	-0.154***	-0.186***	-0.053***	-0.101***	-0.132***	-0.246***
	(0.018)	(0.020)	(0.046)	(0.052)	(0.014)	(0.021)	(0.030)	(0.045)
Observations	814388	501376	814388	501376	2173127	1011399	2173127	1011399
# of Clusters	3412	2467	3412	2467	3863	2764	3863	2764
$\beta 1 + \beta 2$	-0.023	-0.023	-0.049	-0.035	-0.020	-0.037	-0.042	-0.079
Pvalue	0.045	0.070	0.111	0.333	0.029	0.003	0.042	0.006
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. In columns 1 through 4 observations are weighting by the inverse probability score, and in columns 5 through 8 treatment status is assigned by propensity score bins. See text for additional details. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7: Sensitivity to Partnership Date Definition

	Partnership on March 19th				Excluding Intermediate Period			
	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue	(5) Orders	(6) Orders	(7) Revenue	(8) Revenue
Treat*Post	0.064***	0.078***	0.150***	0.190***	0.058***	0.074***	0.138***	0.184***
	(0.010)	(0.012)	(0.025)	(0.031)	(0.011)	(0.013)	(0.027)	(0.032)
Treat*Post*Low	-0.114***	-0.136***	-0.231***	-0.287***	-0.112***	-0.133***	-0.228***	-0.287***
	(0.013)	(0.016)	(0.033)	(0.040)	(0.014)	(0.016)	(0.035)	(0.041)
Observations	2068197	1257612	2068197	1257612	2040166	1231415	2040166	1231415
# of Clusters	3862	2714	3862	2714	3832	2723	3832	2723
$\beta1 + \beta2$	-0.050	-0.058	-0.080	-0.096	-0.054	-0.059	-0.091	-0.103
Pvalue	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. In columns 5 through 8 observation between February and March 19th are excluded form the analysis. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. In columns 1 through 4 post is indicator for before and after March 19th. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A8: Sensitivity to Business Attrition

	Exculding (Eventual) Exiters				Exit as Zero Sales & Revenue			
	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue	(5) Orders	(6) Orders	(7) Revenue	(8) Revenue
Treat*Post	0.055***	0.074***	0.125***	0.171***	0.051***	0.067***	0.134***	0.177***
	(0.010)	(0.013)	(0.026)	(0.032)	(0.011)	(0.013)	(0.027)	(0.034)
Treat*Post*Low	-0.099***	-0.123***	-0.194***	-0.246***	-0.096***	-0.113***	-0.201***	-0.239***
	(0.014)	(0.016)	(0.035)	(0.042)	(0.013)	(0.016)	(0.034)	(0.042)
Observations	1914696	1116849	1914696	1116849	2282647	1387335	2282647	1387335
# of Clusters	3582	2509	3582	2509	3872	2722	3872	2722
$\beta 1 + \beta 2$	-0.044	-0.049	-0.069	-0.075	-0.044	-0.046	-0.067	-0.062
Pvalue	0.000	0.000	0.003	0.009	0.000	0.000	0.003	0.020
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. An observation is business-week. Column 1 through 4 excluded all firms that exit YTP during the analysis, in columns 5 through 8 sales and revenue of existing firms are coded as zero instead of missing. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A9: Sensitivity to City-Level Time Trends

	(1) Orders	(2) Orders	(3) Revenue	(4) Revenue
Treat*Post	0.060***	0.067***	0.139***	0.165***
	(0.010)	(0.012)	(0.025)	(0.033)
Treat*Post*Low	-0.102***	-0.126***	-0.217***	-0.274***
	(0.014)	(0.016)	(0.035)	(0.043)
Observations	2173244	1321619	2173244	1321619
# of Clusters	3875	2725	3875	2725
$\beta 1 + \beta 2$	-0.042	-0.059	-0.078	-0.109
Pvalue	0.000	0.000	0.001	0.000
Treatment Def.	Median	25<>75	Median	25<>75
Quality Def.	25<>75	25<>75	25<>75	25<>75

Note: This table reports regression coefficients from 4 separate regressions. An observation is business-week. The dependent variables are the per-business inverse hyperbolic sine transformation of weekly number of orders and weekly-revenue, and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. In addition, the specification allows for city-level time trends in establishment outcome. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A10: Effect of Entry on Users' Type

	Weekly Orders	Weekly Expenditure	Dollar Ratings	Variety (Restautrants)	Variety (Categories)
Treat*Post	-0.004	-0.004	0.002	-0.012***	-0.003
	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
# of Clusters	2842	2843	2810	2843	2843

Note: This table reports regression coefficients from 6 separate regressions. An observation is one user. The sample includes only users in cities that were effected by the partnership and only includes behavior in the period after integration. The dependent variables are the inverse hyperbolic sine transformation of the outcomes indicated in sub-headings, and should be interpreted as percentage changes. Coefficients represent the interaction between a dummy for user joined YTP post implementation and treatment status is an indicator for whether the city experienced an above median change in the percentage of businesses on YTP. All regressions include city fixed effects. Number of observations are not reported in order to protect proprietary data. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A11: Effect of Entry on Incumbents' Subsequent Ratings- Robustness Checks

	Placebo Test				Type of Raters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat*Post	0.003	-0.001	0.008	-0.007	0.001	0.001	0.003	0.002
	(0.008)	(0.010)	(0.009)	(0.015)	(0.003)	(0.003)	(0.003)	(0.004)
Treat*Post*Low			-0.059**	-0.027			-0.008	-0.004
			(0.026)	(0.020)			(0.008)	(0.010)
Observations	834964	488933	393194	233178	1464204	859456	695502	413378
# of Clusters	3356	2243	3110	2095	3827	2655	3680	2569
$\beta 1 + \beta 2$			-0.051	-0.034			-0.005	-0.002
Pvalue			0.050	0.109			0.539	0.810
Treatment Def.	Median	25<>75	Median	25<>75	Median	25<>75	Median	25<>75
Quality Def.			25<>75	25<>75			25<>75	25<>75

Note: This table reports regression coefficients from 8 separate regressions. In columns 1 through 4 an observation is business-week, in columns 5-8 an observation is user who rated a YTP restaurant and. The outcome is described further in the text. The dependent variable is the inverse hyperbolic sine transformation of the average rating received in a given week (columns 1-4) and the average rating given by user (column 5-8), and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A12: Effect of Entry on Advertising Purchases By Expense Type

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Fixed Revenue					
Treat*Post	-0.015	-0.024*	-0.021	-0.031	-0.028	-0.052*
	(0.009)	(0.012)	(0.012)	(0.017)	(0.015)	(0.022)
Treat*Post*Low			0.010	0.041	0.003	0.055*
			(0.018)	(0.022)	(0.022)	(0.028)
$\beta1 + \beta2$			-0.011	0.010	-0.025	0.003
Pvalue			0.422	0.455	0.113	0.836
	Panel B: Variable Revenue					
Treat*Post	-0.031**	-0.040**	-0.037**	-0.034	-0.054**	-0.083**
	(0.013)	(0.017)	(0.018)	(0.027)	(0.023)	(0.034)
Treat*Post*Low			0.009	0.045	0.023	0.097**
			(0.025)	(0.032)	(0.032)	(0.041)
Observations	4409516	2623347	4409516	2173244	2623347	1321619
# of Clusters	3964	2781	3964	3875	2781	2725
$\beta1 + \beta2$			-0.028	0.011	-0.031	0.014
Pvalue			0.135	0.553	0.189	0.537
Treatment Def.	Median	25<>75	Median	Median	25<>75	25<>75
Quality Def.			Median	25<>75	Median	25<>75

Note: This table reports regression coefficients from 12 separate regressions, 6 per panel. An observation is business-week. The dependent variable is the inverse hyperbolic sine transformation of total advertising expenditure on Yelp by expense type and should be interpreted as percentage changes. The sum of the coefficient is presented below each panel along with the corresponding Pvalue. The interaction between post and quality level indicators is omitted for brevity. Treatment status and quality definitions are indicated below the table and are described further in the text. All regressions include business and week-state fixed effects. Standard errors are in parentheses and are clustered at the city level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A13: Simulations Using Demographic Data

	10%	20%	40%	60%	80%	100%
Δ Welfare	0.49	1.19	1.88	2.34	2.57	2.77
Δ Market Size	0.62	1.70	2.95	3.88	4.38	4.79
Δ Sales- By Rating Quantile						
1st Quantile	-0.03	-0.08	-0.21	-0.36	-0.50	-0.63
2nd Quantile	0.08	0.09	-0.07	-0.20	-0.44	-0.41
3rd Quantile	0.25	0.48	0.40	0.07	-0.05	-0.24
4th Quantile	0.29	0.63	0.74	0.51	0.34	0.32
5th Quantile	0.29	0.79	1.31	1.59	1.54	1.36

Note: This table reports the average results of 500,000 simulated markets. The parameters used to simulate the data are presented in column 4 of Table 12. The simulation algorithm is described in appendix A.5. The tables presents the percentage change in outcomes from the baseline. In the baseline, 5% of firms in the market are on the platform. Outcomes are indicated in row names and subheading. Column headers indicate the simulated share of firms on the platform.

A.2 Proofs

Proof of Lemma 1 In this proof (and only this proof) I assume that the idiosyncratic error is distributed according to the extreme value type 1 distribution, which is the distribution used on the structural model. Also, without loss of generality, I present the proof using expected quality distribution. Assuming a symmetric equilibrium, high-quality firms and low-quality firms solve:

$$\begin{aligned}
\max_{p_j} \Pi_j^h &= H(E[max(u)] - \omega) * \frac{e^{q_h - p_j}}{1 + e^{q_h - p_j} + \frac{n-1}{2} e^{q_h - p_h} + \frac{n}{2} e^{-p_l}} (p_j - r) \\
0 &= \frac{2(r - p_h + 1)e^{p_h + p_l + p_j} + n(r - p_h + 1)e^{p_h + p_j} - (n-1)(-r + p_h - 1)e^{p_l + q_h + p_j} + 2e^{p_h + p_l + q_h}}{(2e^{p_h + p_l + q_h} + 2e^{p_h + p_l + p_j} + ne^{p_h + p_j} + (n-2)e^{p_l + q_h + p_j})^2} \\
0 &= 2(r - p_h + 1)e^{p_h + p_l + p_h} + n(r - p_h + 1)e^{p_h + p_h} - (n-1)(-r + p_h - 1)e^{p_l + q_h + p_h} + 2e^{p_h + p_l + q_h} \\
0 &= 2Ae^{p_h + p_l} + [(n-1)A - 2]e^{p_l + q_h} + nAe^{p_h} \qquad \text{(A.1)}
\end{aligned}$$

$$\begin{aligned}
\max_{p_j} \Pi_j^l &= H(E[max(u)] - \omega) * \frac{e^{-p_j}}{1 + e^{-p_j} + \frac{n-1}{2} e^{-p_l} + \frac{n}{2} e^{q_h - p_h}} (p_j - r) \\
0 &= \frac{2(-r + p_j - 1)e^{p_h + p_l + p_j} + (n-1)(-c + p_j - 1)e^{p_h + p_j} + n(-r + p_j - 1)e^{p_l + q_h + p_j} - 2e^{p_h + p_l}}{(2e^{p_h + p_l + p_j} + 2e^{p_h + p_l} + (n-2)e^{p_h + p_j} + ne^{p_l + q_h + p_j})^2} \\
0 &= 2(-r + p_j - 1)e^{p_h + p_l + p_l} + (n-1)(-r + p_j - 1)e^{p_h + p_l} + n(-r + p_j - 1)e^{p_l + q_h + p_l} - 2e^{p_h + p_l} \\
0 &= 2Be^{p_h + p_l} + [(n-1)B - 2]e^{p_h} + nBe^{p_l + q_h} \qquad \text{(A.2)}
\end{aligned}$$

Where $A = (p_h - r - 1)$ and $B = (p_l - r - 1)$.

Combine the A.1 and A.2 to get:

$$\begin{aligned}
[A((n-1)B - 2) - nAB]e^{p_h} &= -[nBA - B((n-1)A - 2)]e^{p_l + q_h} \\
e^{p_h} &= \frac{B(A+2)}{A(B+2)} e^{p_l + q_h} \\
e^{p_h} &= \delta e^{p_l + q_h} \qquad \text{(A.3)}
\end{aligned}$$

Substitute e^{p_h} with $\delta e^{p_l + q_h}$ into A.2 to get:

$$e^{p_l} = -[\frac{n-1}{2} - \frac{1}{B} + \frac{n}{2\delta}]$$

Since $\delta = e^{p_h - p_l - q_h} > 0$, $\frac{n}{2\delta} > 0$, and $\frac{n-1}{2} > 0$, it must be that $B > 0 \implies p_l > r + 1$. Similarly, one can show that $p_h > r + 1$ by plugging A.3 into A.1.

Now, assume towards contradiction that $p_h = p_l$. This implies that $A = B$ and $\delta = \frac{B(A+2)}{A(B+2)} = 1$. But then by Equation A.3 $e^{p_h} = e^{p_h + q_h} \implies q_h = 0$, a contradiction.

Now, assume toward contradiction the $B > A$. Then $\delta = \dfrac{B(A+2)}{A(B+2)} > 1$. However, noting that $B > A \implies p_l > p_h$ together with Equation A.3: $ln(A) = p_h - p_l - q_h$ implies that $0 > p_h - p_l - q_h = ln(A) \implies A < 1$, a contradiction. Therefore $p_l < p_h$ implying that $A < 1 \implies ln(A) < 0 \implies p_h - p_l - q_h < 0 \implies q_h - p_h > -p_l$.

Proof of Lemma 2 Inspecting equations 1.1 and 1.2, the first element does not depend of firm's quality. In contrast, lemma 1 implies that the second term, the per-unit profit, is larger for high-quality firms compared to low-quality firms. The third term captures the number of sales: The first element is equivalent for high-quality and low-quality, and the second and third elements are larger for high-types since $p_h - q_h < p_l$. Thus, high-quality firms sell more, charge higher prices, and generate higher profits.

Proof of Proposition 1 The following proof shows the effect of adding a high-quality firm. The proof when adding a low-quality firm is similar. Also, I assume here that N is sufficiently large. It is simpler to study the derivative of the natural logarithm, $\frac{\partial ln(pi)}{\partial N}$:

$$\begin{aligned} ln(\pi_h) &= ln(H(E_\epsilon[max(u_{ij})] > \omega + c_i)) + ln(p_h - r) + \\ &\quad N_h ln(1 - \tilde{G}(0)) + N_l ln(1 - \tilde{G}(\bar{U_l} - \bar{U_h})) + ln(1 - G(\omega - \bar{U_h})) \qquad (A.4) \\ ln(\pi_l) &= ln(H(E_\epsilon[max(u_{ij})] > \omega + c_i)) + ln(p_l - r) + \\ &\quad N_h ln(1 - \tilde{G}(\bar{U_h} - \bar{U_l})) + N_l ln(1 - \tilde{G}(0)) + ln(1 - G(\omega - \bar{U_l})) \qquad (A.5) \end{aligned}$$

Where $\bar{U_j} \equiv q_j - p_j$ for $j \in \{h, l\}$. By the envelope theorem, we can ignore the impact the N_h has on p_h, p_l, $U_h(p_h)$, and $U_l(p_l)$, which means that we can ignore the second, fourth, and fifth terms in equations A.4 and A.5.[1] The third term in equation A.5 is more negative than the third term in equation A.4, and both are negative:[2]

$$ln(1 - \tilde{G}(\bar{U_h} - \bar{U_l})) < ln(1 - \tilde{G}(0)) < 0 \text{ since } \bar{U_h} > \bar{U_l}$$

Since the effect of the first term $\frac{\partial ln(H(E_\epsilon[max(u_{ij})] > \omega + c_i))}{\partial N_h}$ is the same in on both types, this proved the second part of proposition 1.

Finally, we have to show that the first term is increasing in N_h. Since the both $ln(\cdot)$ and $H(\cdot)$ are monotonically increasing transformations, and $\omega + c_i$ are independent of N_h, we only have to show that $E_\epsilon[max(u_{ij})]$ is increasing in N_h. This follows from the fact that the maximum is increasing in the number of draws.[3] Together, this implies that the total effects on $ln(\pi_h)$ and $ln(\pi_l)$ are ambiguous, which established the first part of the proposition.

[1]This depends on the assumption that N is large is quite large, so that the change in N_h is comparably small. Alternatively, we can reconstruct the model to have a continuum of firms with a measure μ and explore the impact of an infinitesimal increase in μ. For this formulation, however, we would have to make a direct assumption that $u'_\mu > 0$, at least in expectation, i.e., that consumers have some love-for-variety (which could be motivated by the model above).

[2]The same would be true if we were considering an increase in N_l:

$$ln(1 - \tilde{G}(0)) < ln(1 - \tilde{G}(\bar{U_l} - \bar{U_h})) < 0 \text{ since } \bar{U_h} > \bar{U_l}$$

[3]A simple, yet not very elegant proof, denote $max(x_1, ...x_m)$ by a and the pdf of a as $g(a)$, then:

$$\begin{aligned} E[max(x_1, ..., x_m, x_{m+1})] &= \int \{P(x_{m+1} > a)E[x_{m+1}|x_{m+1} > a] + (1 - P(x_{m+1} > a))a\}g(a)da \\ &= \int \tilde{a}g(a)da > \int ag(a)da = E(max(x_1, ...x_m)) \end{aligned}$$

A.3 Data Appendix

Sample selection and main results While Yelp keeps data on when restaurants join and exit YTP, I found multiple cases where transactions were made prior to a business 'entering' the platform or after the business 'exited' the platform. In cases on inconsistent data, I always code entry as the earliest date of the two and exit the the later date. For this reason, I leave a margin for 8 weeks at the beginning and end of my sample to separate between businesses with zero sales to businesses that have left the platform. Weeks in which I do not observe any transactions are coded as zero orders and zero revenue. Sales and revenue are coded as missing for the week before entry or after exit.[4] The final data used in the analysis consist of 88 weeks, from March 2017 to December 2018. I limit the analysis to cities in which there are ten or more businesses on the standard Yelp platform, since in very small places treatment intensities are extremely large mechanically. I excluded businesses that are marked by Yelp as bogus, spammy, or that are removed from users' search results. 310 businesses in 291 cities are dropped from the analysis which amount to less than 30,000 observations in my data. The final sample consists of 3,956 cities. For the main part for the analysis, I use only the incumbent businesses, which joined YTP prior to the partnership with Grubhub; there are a total of 56,493 incumbent businesses and over 4 million business-week observations.

The Yelp system does not store historic businesses' star-ratings. To calculate businesses' rating on the eve of integration, I take the mean over all preceding review. I exclude reviews that are marked by Yelp as untrustworthy, or are removed from business' page. To test whether this is a good approximation, I use the same method to calculate the current Yelp rating and find that there is a correlation of over 0.95. Restaurant categories are based on Yelp's classification. For the robustness checks presented in appendix A.4.2, I include only the top 21 most prevalent food categories (out of 244), which include 87% of all observations in the sample. Generally, Yelp collects little demographic information on its users. Users are encouraged to enter their gender and date of birth, but I found the fields to be mostly missing in my sample. Thus, I do not use individual-level characteristics in my analysis. For each user, however, I do observe the full history of transactions on YTP, which is used to differentiate between new and repeating consumers.

Prices As mentioned in Section 1.5.4 price data is extremely problematic since the data only includes prices for ordered items, as opposed to menu prices, and not all dish modifications are recorded. To address this issue, I attempt to identify 'true' prices using an algorithm developed for Luca & Reshef (in writing). The algorithm takes several step: (1) drop all item for which the name suggests possible modifications (e.g., 'customize', 'create' etc.) as well as items that were discounted or orders when a coupon was used, (2) Include only prices that appear 3 or more times, but no less than 10% of total price observations for the item, (3) calculate the inter quantile range (in terms of time) for each price level and excluded observations that are above (below) 1.5 times the upper (lower) bound, (4) find the first and last occurrence of each item-price, and (5) use only prices that do not overlap. This algorithm is extremely restrictive and ultimately discards more than half of the food items in the sample.

[4]Appendix A.4.2 tests the sensitivity of the results to precise definitions.

Search Orders data and search data are handled by different parts of the organization, and more importantly, are stored in different data clusters. Consequently, joining the two datasets is not a trivial task. To identify the search sessions which lead to an order, I develop an algorithm that matches each order with the most recent search session conducted by the user prior to finalizing the order. The algorithm has several disadvantages: First, it will not be able to match an order to a search session if the user was not signed in during the search process. Second, when a user performs multiple searches on the same day, the algorithm only picks up the last session. This might be an issue if users use multiple search sessions to choose a restaurant to order from. Though these issues create additional noise and reduce statistical power, they are unlikely to bias the results in any particular direction.

Advertising Some additional institutional details that are omitted from the text for brevity: Profile enhancement is a bundle designed to increase the attractiveness of Yelp business page. It includes several upgrades such as slideshow and videos, access to Yelp account management support and data analysis tools, and removal of ads purchased by competitors the Yelp business page. Advertisements on Yelp appear in three places: (1) Ads appear at a premium location, above the organic Yelp search results, for keywords that are related to the business. (2) Ads appear on competitors' businesses pages (only if they are not advertising on Yelp as well) (3) On the Yelp mobile app. Payment is based, for the most part, on cost per click (CPC). Bidding for ads on Yelp is done using Yelp's auto-bidding algorithm; the only choice the client has is the period advertisement budget, which is usually exhausted. For the most cases, profile enhancement and targeted ads are purchased in one package. Some, more sophisticated, businesses purchase a custom or a la carte services. This type is both very rare and not well documented, so I exclude it form my analysis. Yelp uses a complicated algorithm to determine the payment per ads which includes user clicks, traffic, and actions on the business page. The Yelp auto-bidding algorithm cannot be manually overridden by the client. The algorithm determines key words to bid on, as well as bid amounts. I use data on business-level weekly revenue collected by Yelp to estimate the effect of entry on advertising behavior. I restrict attention to campaigns purchased at the local-level, excluding national- or franchise-level campaigns.

Placebo tests To test the validity of the research design, I consider three outcomes: First, the weekly flow for new businesses that are classified by Yelp as either 'food' or 'restaurants'. Second, the weekly flow of new ratings per business, and the average rating given. I include only businesses that are classified as either 'food' or 'restaurants' and exclude businesses that are marked as bogus, spammy, or that are removed from users' search. Importantly, businesses participating in YTP are also excluded form the analysis. I also exclude review that are marked as untrustworthy, were removed by Yelp, were given by paid users, or that are given by consumers that use YTP. To test alternative explanations for the increase in ratings, I test whether selection of more lenient reviewer is driving the results.[5] To test for reviewer leniency, I construct the leave out average of all reviews given by a user. Weekly ranking in search order results is the average weekly rank across all search results in which the business appears. Businesses that are marked as bogus, spammy, or that are removed from users' search, and review that are marked as untrustworthy, were removed

[5] I also test a specification in which the implementation date is countfactually set at the middle of the pre-period. The results are reported in appendix A.4.3.

by Yelp, or were given by paid users, are excluded. Naturally, user who only gave one review in total, are excluded as well.

Data For Structural Model Demographic information on population, gender, age and income comes from the American Community Survey 5-Year Data 2017. Total population in each county is divided by 3, to approximate the number of households. I experimented with several factors ranging from 2 to 4, and it appears the main qualitative results are robust to such changes. Age is binned into 10-24, 25-44, and the omitted category is 45 and above. Adjusted gross income is binned by annual household income of (in thousands): below \$50, \$50-\$100, and \$100-\$200. These data are available at the county level and are merged with the city-level data by name and state name combination. Multiple matches are determined randomly. Data are merged with the main dataset by 5-digit zip codes. CSA data are merged by city-state name and county-state fips code (when available) combination. Beside, the county-level income data, I also use data on income at the zip-code level is used as part of businesses attributes. These data come from the IRS Individual Income Tax ZIP Code Data 2016. I use a continuous measure of the 6 Adjusted Gross Revenue bins presented above.

In order to reduce computation burden and since many businesses do not any orders at a given week, I aggregate the data by city and 20 weeks bins. The data are trimmed to include only 4 bins for each city, two in the period before integration and two afterwards. The outcome variable is the total number of order per business during each 20 weeks period. Since the Dollar Ratings for a given business rarely change over time, I use the same rating for a given business in all four markets. Rating of businesses are calculated by the rating at the mid-point of each 20 weeks period. Ranking and ranking within category are defined by the relative ranking of the businesses in a city-period or city-period-type combination. Total ranking in the city take values of 1 to 5 by the ranking quantile in the city. The reason to use the ranking quantile as opposed to absolute rank is that, for the simulations, the number of businesses increases and so does the range of ranks. Mechanically, at very large size the rank number can increase by hundreds and even thousands of percents and renders all other covariates meaningless. For the decomposition by rating level, I use rating quantiles across all cities in the sample. I aggregate food categories to 10 main categories.[6] To improve simulation performance, the final sample excludes markets with less than 5 businesses (less than 5% of the sample) or more than 250 (less than 1% of cities). Finally, since the discrete choice models I use do not allow for zero market shares, I correct for zero market share by adding one order to that business in that city-period time.

To get the share of users of the platform out of the total number of potential buyer, I use the search data. I define user interested in the platform as the number of unique user IDs in a given city-period combination who have search for variations of the words “delivery”, “takeout” or “Pickup”, or use any of the YTP filters. There are some limitations to this definition: First, while I am interested solely in food orders, this algorithm will also pick up individuals who are interested in, for instance, flower delivery. Second, if users conduct search without logging in, the system will document them as new user. Both of these types of issues are likely to results in overestimation of the number of users on the platforms.

[6]The ten categories are: Asian, pizza, Mexican & Latin, European, Arab-Indian, meat & seafood, American, coffee & pastry, sandwiches, and other.

Accordingly, the mean share of platform users across markets is 25%, which seems excessive. Nevertheless, though this is an inaccurate measure, it is unlikely to be correlated with any of the mechanisms of interest and thus I would does not change the qualitative nature of the results.

A.4 Robustness Checks Appendix

This Appendix discusses the robustness checks conducted for the main and subsequent results. All relevant figures and tables can be found in Appendix A.1. Subsection A.4.1 presents tests for the validity of the research design described in Section 1.4.2. Subsection A.4.2 presents the robustness checks for Section 1.5.3, robustness of the main results. Finally, subsection A.4.3 presents the robustness checks for the investment in quality results presented in Section 1.5.4.

A.4.1 Validity of the Research Design

This section presents additional tests to support the parallel trend assumption, which is the key identifying assumption of the difference-in-differences research design. First, I test for differences in trends between treated and control cities in the period before the partnership with Grubhub, and second, I test for difference in trends on outcomes that are unlikely to be affected by the institutional change. The research design is presented in Section 1.4.2.

Pre-trends The first suggestive evidence of parallel trends absent of treatment is to examine whether the main outcomes of interest trend similarly in the prior to the Grubhub partnership. Graphic evidence are presented and discussed in Section 1.5. In Table A1, I present a formal placebo test in which I counterfactually set the integration date to the middle of the pre-treatment period. I do not find any significant effects of the placebo on the average effect (column 1) or when examining the effect on high- or low-quality firms separately (columns 2 and 3). This results suggests that the main results are not driven by initial differences in trends between treated and control cities.

Placebo on non-YTP outcomes A second potential concern is the break in trends is driven by other unobserved changes at the city level that are unrelated to the Grubhub partnership. If that is indeed the case, then we can expect to find significant differences in other city-level outcomes, not directly related to food ordering. I conduct several placebo tests to examine whether the partnership is correlated with outcomes of *non-YTP* businesses, such as the number of businesses on Yelp, the restaurant average weekly ratings, and the number of new weekly reviews per business. Table A2 presents the results. Panel A presents the results of estimating equation 1.4 the percentage change in the number of new restaurants on Yelp as the outcome variable. I do not find any significant effects, which suggests that, in general, the food industry is growing similarly in treated cities and control cities. Due to the large number of observations, Panels B and C present results at the monthly level Again, columns 1 through 3 do not find any significant effects of treatment on the average weekly rating, or the number of review per business for *non-YTP restaurants*. Similarly, columns 4 and 5 present null effects of the partnership by quality levels.[7] Taken together,

[7]In contrast, in section 1.5.4 I find that treatment does effect the average and high-quality weekly ratings

the null findings suggest that the results are not driven by unobserved changes in the city, the restaurant industry, or Yelp usage.

A.4.2 Robustness of the Main Results

Market definition-geographic area Table A3 presents the estimation results using alternative geographical definitions for the relevant market. Note that the number of observations decreases since not all observations include zip-code and county data. Columns 1-4 and 5-8 present the results when using the 5-digit code and county as the relevant markets, respectively. The first two columns in each group show the effect on number of orders and the last two columns in each group show the effect on weekly revenue. Qualitatively, the results are similar to the main estimation results: Entry leads to more sales and higher revenue for high-quality restaurants, and vice versa for low quality businesses. The point-estimates of the effects vary across specifications and market definition. This is especially true for when estimating the effect on revenue, using the county as the relevant market due to the low number of clusters and the noisiness of weekly revenue. Nevertheless, estimates are centered around the main results, do not change signs, and are generally statistically significant. For instance, columns 2 and 6 estimate the effect of entry on weekly orders using the sharp definitions for treatment and rating. They find a treatment effect of 4.9% to 6.6% for high types and -6.2% to -4.2% for low types. In comparison, using the city as the relevant market, I find an effect of 6.5% for high types and -5.4% for low types (column 2 in Table 4).

Market definition-food category An alternative definition treats each food category (pizza, Chinese, Mexican etc.) and city combination as a separate market. There are a few important disadvantages to this definition: First, it imposes strong restrictions on consumers' decision-making process. In particular, this market definition implicitly assumes that consumers first decide which type of food they want to eat and only then choose the particular restaurant. This assumption is violated if, for instance, consumers search all restaurants in their area and choose the highest rated one. Second, even if restaurants only compete within food category, there are likely to be positive spillovers across categories, i.e., the SUTVA assumption is unlikely to hold for the market size effect. To see why this is the case, consider the implications of a market expansion in only one food category; the definition above implies that this will have no spillovers to other food types, and that consumers in that city will keep purchasing other food types from alternative channels. This assumption is extremely restrictive and is likely to fail. Finally, the borders of "food categories" are not clearly defined. There are over 244 unique food categories in the data, and some, such as Japanese food and sushi, are clearly not mutually exclusive and are likely to be decent substitutes. To address this issue without taking a stand on food "similarities", I restrict my analysis to the 21 most prevalent food types (food categories with more than 50,000 observations), which constitute 87% of all observations.[8] I then reconstruct the entire dataset using the city-category market definition. I reapply the same rules as the main analysis, including redefining treatment intensities and high and low quality firms.

of YTP restaurants in treated cities compared to untreated cities.

[8]These categories are (in the order of importance): pizza, Chinese, sandwiches, Mexican, traditional American, Japanese, Indpak, hotdogs, Thai, Mediterranean, Italian, breakfast and brunch, seafood, cafes, burgers, new American, barbecue, delis, Asian fusion, and Vietnamese.

The results are presented in the first part of Table A4. First of all, note that the number of observations decrease since we restrict attention to a subset of restaurants, and since, as in the main analysis, very small markets are excluded from the analysis. In contrast, the number of clusters substantially increase because now each city is separated into several markets. The main qualitative results are robust to the change in market definition; I find statistically significant positive effect on weekly sales and revenue for high-quality firms, and vice versa for low-quality firms. The effects on low firms are about the same magnitude for low-quality firms, averaging around –4.5% and –8% for sales and revenue, respectively. The effects on high firms are smaller than the main specification. For instance, the estimated effect on high-quality restaurants' weekly number of orders is only 2.7%–3% compared to 5%–6.5% in the main analysis. This result is consistent with arguments regarding positive spillovers across "markets" which introduce attenuation bias to the estimated coefficient.

Treatment definition The second part of Table A4 presents the results using the alternative market definition. This treatment definition is substantially noisier, especially due to places that had only a handful of restaurants on YTP prior to the integration. For instance, by that definition, some markets have a treatment intensity of over 200%. Nevertheless, the estimates of the effect of entry on weekly sales and revenue are robust to the alternative definition of treatment intensity; all coefficients have the same size and similar magnitudes as the main analysis, though the revenue estimates are substantially noisier. For example, column 6 suggests that, using the sharp definitions of both treatment and ratings, entry increases weekly sales of high-quality restaurants by 5.4% and decreases weekly sales by 4.6% compared to 6.5% and -5.4% when using the standard treatment definition.

Outliers To test whether the main results are driven by outliers, I perform two separate robustness checks. The results are presented in Table A5. First, I exclude the cities that are in the top (bottom) 5% in terms of the number of businesses on YTP prior to integration. This exercise turns out to be quite restrictive: To begin with, many small towns had only a handful of businesses prior to integration, so there is substantial mass at 5%. Additionally, the largest cities, with the most incumbent businesses, naturally contribute the most observations to the analysis. Thus, excluding outliers reduces both the number of clusters and the number of observations sententially. As columns 1-4 show, the qualitative results are similar to the main specification and are statistically significant. The estimated size of the effects diminishes in comparison to the main analysis, but we cannot reject the null hypothesis that the estimates are the same as the main specification.

Second, I estimate p-values using randomization inference instead of a traditional sampling-based approach. Randomization inference performs better in settings with concentration of leverage, the degree to which individual observations of right-hand side variables take extreme values and are influential, in a few observations (Young, 2016). Due to the large number of observations and the time it takes to run these specification I perform only 1000 iterations for each specifications. All of the p-values on the coefficients of interest are zero, i.e., the estimated effects were larger than all of the 1000 randomized treatment effects. Taking these results together, I conclude that it is unlikely that the estimates are driven solely by outliers.

Initial differences between treatment and control Though there are similar trends in treated and control markets prior to integration, there are substantial differences between markets. Specifically, treated markets tend to have more restaurants, have more restaurants on YTP, and have a larger share of restaurants on YTP.

To address the concern that these initial differences are driving the results, I perform two robustness checks: Firstly, I use inverse propensity score weighting to correct for the bias (Hirano et al., 2003). I estimate the propensity score using a third-order polynomial logit model with the total number of businesses, total number of businesses on YTP pre-integration, and the share of businesses on YTP pre-integration as predictors. Though there is generally common support on the full interval, I trim propensity above and below the 90th and 10th percentile to correct for differences in mass. The results are presented in Table A6 columns 1 through 4. Though the point estimates are slightly lower, the estimated effect of market expansion on high-quality firms remains positive and statistically significant. The differences between the effects of high- compared to low- quality firms are both economically and statistically significant. The total effects on low-quality firms are both smaller in magnitude and noisier than the main specification

Secondly, since treatment and control looks very different on observables, I conduct an additional analysis, which takes advantage of the fact that treatment intensity is continuous. First, I run a linear probability model of treatment intensity (change in share of businesses on YTP) on a third-order polynomial with the total number of businesses, total number of businesses on YTP pre-integration, and the share of businesses on YTP pre-integration as predictors. Figure A3a presents the distribution of *predicted* treatment intensity by treatment status, where treatment is defined by the median (actual) treatment intensity. City characteristics clearly predict treatment intensity, though there is substantial overlap between the two groups. Secondly, I divide markets into 20 bins based on predicted treatment intensity, with an equal number of markets within each bin. I then assign markets into treatment and control based on their relative treatment intensity in their respective bin. Intuitively, bins with higher predicted intensity will tend to have higher thresholds to be included as treatment. Accordingly, some markets with very high predicted intensity might be coded as control even though their true intensity is relatively high, and vice versa for low intensity markets. Figure A3b presents the distribution of predicted treatment intensity by treatment status, where treatment is defined by the median *within bin* treatment intensity. It is clear the the distribution of treated and control markets is much more similar than in the original sample.

The estimation results, using the new treatment definition, are presented in Table A6 columns 5-8. The point estimates on all coefficients are similar in magnitude to the main specification and are statistically significant. Taken together, these results suggest that initial differences in market characteristics are not driving the main effects.

Additional robustness tests and results As discussed in Section 1.3 the beginning of the partnership between YTP and Grubhub is coded as February 19th, one month before system integration was completed. To test whether the main results are sensitive to this definition, I conduct two additional robustness checks: First, I define the integration date as March the 19th, and second, I completely exclude the period between February to March 19th. The results are presented in Table A7. In general, the main estimates remain statistically significant and are larger in magnitude when using either of these alternative definitions.

Another potential issue is attrition. Approximately 15% of the incumbent business (9,103) leave YTP before the end of period. In the main analysis, they are used as part of the analysis up to their last week before they leave the platform, then they are excluded from analysis. Since attrition is non-random, one potential concern is that using the data this way is biasing the main estimates. To address this concern I show that the main results are robust to two alternative specifications: First, excluding any establishment that ultimately leaves the platform from the analysis all together, and second, coding leaving businesses' weekly number of orders and weekly revenue as zero after existing the platform. The results are presented in Table A8. In general, I find that all of the main estimates remain statistically significant at 1 percent level and have similar magnitude in both of these alternative specifications.

In addition, I test whether new consumers joining the platform after market expansion are different than existing consumers. In particular, I examine for difference in behavior on the platform between new and old consumers in cities that were affected by the partnership. Table A10 presents the results.[9] In general, I find null difference in weekly number of order, expenditure, and the type and variety of restaurants ordered from. The only significant difference is that experienced users order, one average, for a slightly larger variety of restaurants. While statistically significant, the effect is relatively small, 1.2%.

A.4.3 Investment in Quality- Alternative Explanations

In section 1.5.4. I find that entry increased subsequent investments in quality, as measured by Yelp Ratings. This section addresses the alternative explanations to explain this result.

Rating inflation To address the concern that results are driven by rating inflation (Horton and Golden, 2018, Nosko and Tadelis, 2018), columns 1-4 in Table A11 presents a placebo specification in which integration is counterfactually coded at the middle of the pre-treatment period. Columns 1 and 2 find null average effects which hare are both statistically and economically insignificant. Column 3-4 decompose the average effects by quality type; there are null effects for high-rated firms, and marginally significant *negative* effects on low-type. Taken together these results suggest that differential trends in rating inflation are not driving the increases in ratings following integration.

User selection The second concern is that selection into specific services is correlated with rating behavior (Kovács and Sharkey, 2014). For instance, if users who use delivery services also tend to rate more leniently, then increases in online ordering might mechanically drive up the ratings of restaurants. To alleviate this concern, I test for differential changes in raters' leniency. The results are presented in Panel B of Table A11. For each review rating, I calculate the average rating given by the user through her activity on Yelp. I do not find any evidence that raters' leniency changes following integration: The average effects and effects by quality-type on leniency are statistically and economically insignificant under all specifications.

[9] Since individual users are the unit of observation, the exact number of observations is omitted to protect proprietary Yelp data. Generally, the results presented in the table represent the behavior of millions of users. Also, the analysis excludes users that made only one order on the platform.

A.5 Structural Model

Simulations After estimating the fundamentals of the model, I simulate firm performance under different market conditions. I divide all firms into five bins, based on their rating quantile across the whole sample. The simulation goes as follows:

1. I choose the share of firms, P, on the platform out of the total number of restaurants in the city, from 1% to 100%.

2. I then draw one restaurant, X, from the relevant rating quantile.

3. Multiplying P by the total number of businesses in restaurant X's city, gives the number of restaurants in the simulated market, N_s. I then draw $N_s - 1$ additional restaurants from the same CSA. These restaurants consist X's simulated market. Note that the restaurants in X's simulated market are drawn from *all* the restaurant in that CSA, not necessarily from the same rating quantile.

4. I recalculate the ranking and ranking within category for all restaurants in the simulated market, since these depend on the specific set of competitors in the market.[10]

5. I then calculate firm X's market share and the total expected utility from the market. When using demographics shifters on the random coefficient, market shares have to be integrated over the population. In practice, I only take into account the random coefficient on income, and since it discrete I simply take the weighted average over the three income levels.

6. I then repeat step 2-5 a thousand times, choosing different X firms and simulating different markets.

In total, I simulate 500,000 markets (5 rating quantiles, 100 share level for each, 1000 iterations).

Adding Demographic Shifters In Section 1.6, the main specification uses does not include demographic shifters of the random coefficient. The reason for ignoring the impact of observable demographics on the disutility from price level is that all demographics have a statistically insignificant effect on the dollar rating coefficient. In this section, I show that the main qualitative results are robust to the inclusion of demographics, but generally perform worse in fitting the data. Note that I only include income as a shifter, and average over the other demographic characteristics.

Simulation results are presented in Table A13. The first two rows show that the estimated market size effect is even stronger in the full model, both welfare and market size increase more rapidly when including demographic Similarly, firms at all rating quantiles and all market sizes perform better compared to the main specification. These effects are also noticeable in graph A4. In particular, the bliss points are about 20 percentage points higher than in the main specification: 68, 35, 27, and 15 for the fifth, fourth, third and second rating quantiles, respectively. Nevertheless, the main qualitative results hold– high-quality

[10]Since I find null effects on firms' pricing behavior, simulation abstain from changes in pricing.

firms perform strictly better, market size in increasing at a decreasing rate, and sales evolve non-monotonically with sellers participation.

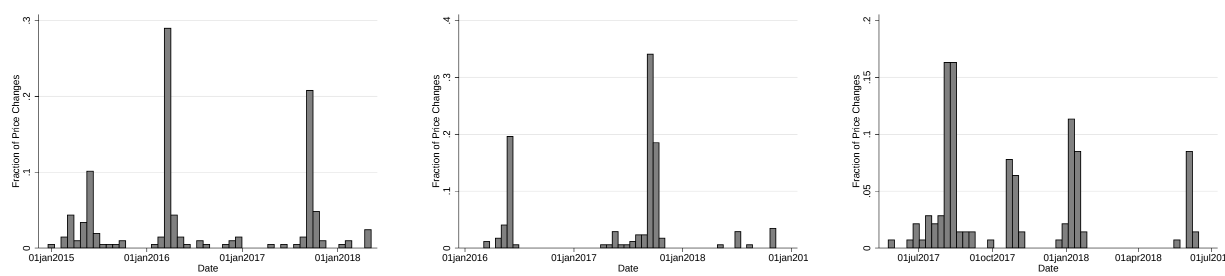
Nevertheless, this set of simulations does poorly in fitting the data. First, the simulated entry cost distribution does a worse job fitting the empirical data, compared to the main specification (not shown). Second, the simulation implies the market size increase of over 45%, while the reduce form result suggest an increase of only 36%. Third, the simulation predicts a loss of only 1% of sales for the lowest rating firms (compared to 4.8% in the reduced form results), and an increase of over 13% percent for the highest rated firms (compared to over 13% in the reduced form results). Together, this results imply that using the point-estimates of the effect of observable demographics on the disutility from price does not do a good job fitting the data, and that those coefficient are better interpreted as zeros.

Appendix B

The Impact of Prices on Firm Reputation

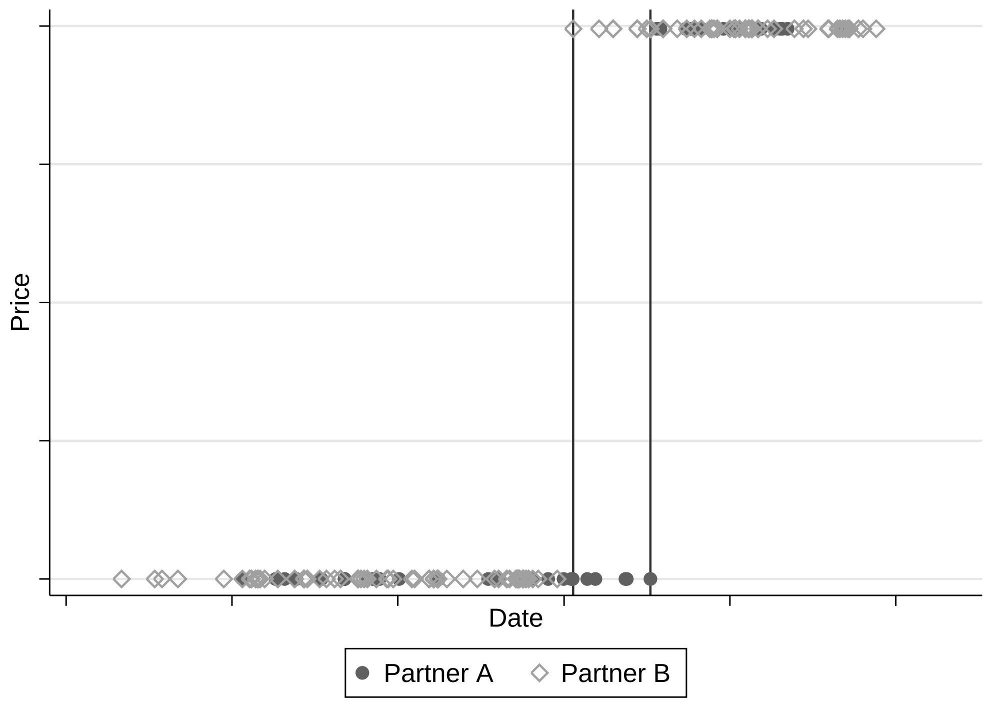
B.1 Appendix Figures and Tables

Figure A1: Visual Representation of Restaurants' Price Changing Patterns



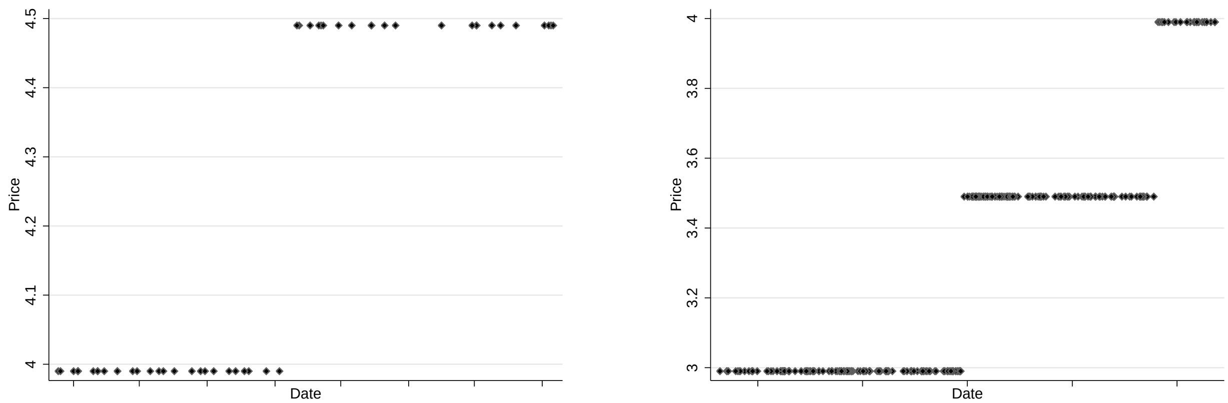
Note: This figure presents the temporal pattern in weekly price changes within restaurant, for the three restaurants with the most price changes in the sample.

Figure A2: Variation in Item-Level Prices Across Delivery Partners

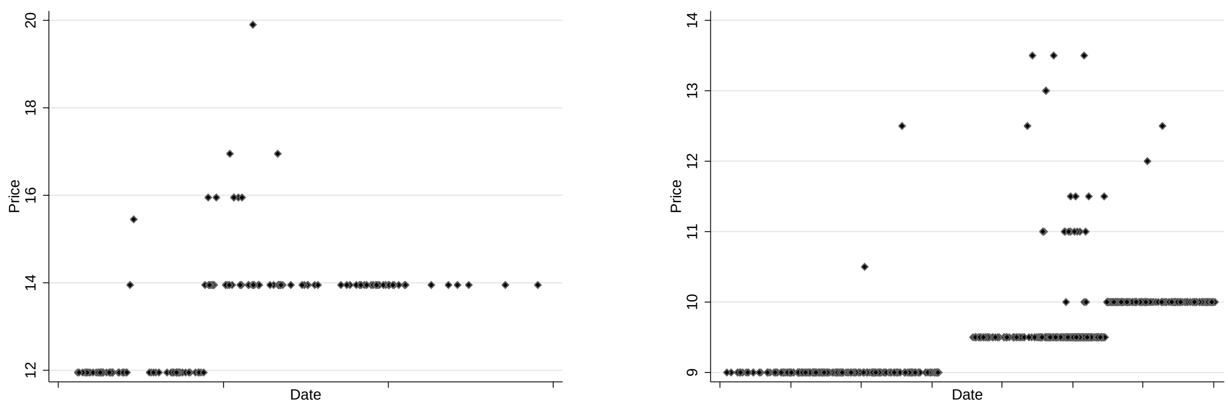


Note: This figure presents the price of one menu item across two delivery partners, marked in A and B.

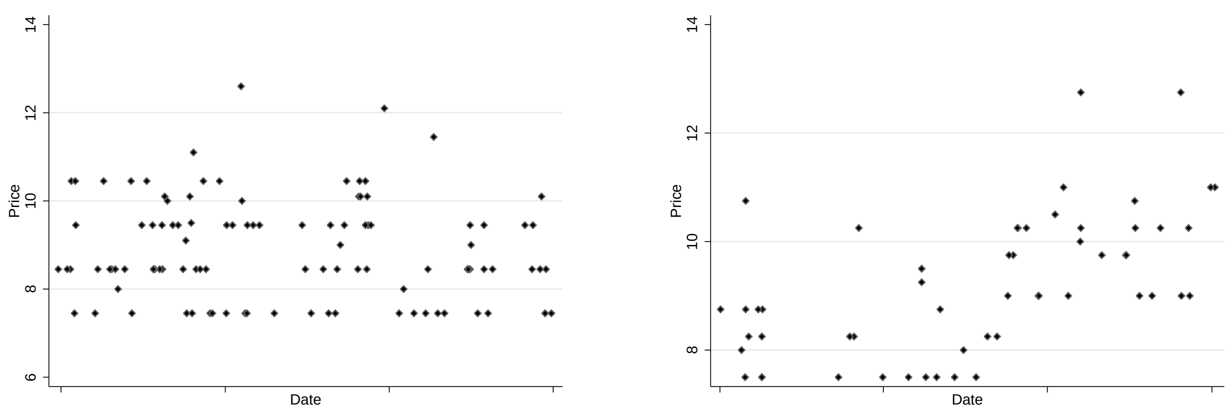
Figure A3: Examples of Item-Level Price Variation Over Time



(a) "Good" price variation



(b) "Good" price variation with noise



(c) "Bad" price variation

Note: This figure presents the temporal pattern in weekly price changes for six different items.

Table A1: The Correlation Between Dollar-Rating and Star-Rating

	(1)	(2)	(3)	(4)	(5)	(6)
$$	-0.025*	-0.025*	0.014	0.085	-0.035	0.021
	(0.013)	(0.013)	(0.012)	(0.063)	(0.023)	(0.029)
$$$	0.078***	0.078***	0.112***	0.228***	0.064*	0.125***
	(0.020)	(0.020)	(0.019)	(0.065)	(0.033)	(0.035)
$$$$	0.139***	0.139***	0.173***	0.230**	0.110	0.147*
	(0.042)	(0.042)	(0.043)	(0.092)	(0.081)	(0.086)
Observations	31663	31663	34218	35162	35133	35162
Controls	Zip	Zip X Cat	Zip X Type	City	City X Cat	City X Type
Clusters	Zip X Cat	Zip X Cat	Zip X Cat	City X Cat	City X Cat	City X Cat

Note: This table reports regression coefficients from six separate regressions. An observation is a (rated) restaurant. The sample includes all restaurants in New York city, Los Angeles, and Houston on January 2019. Outcome is the Yelp Star Rating and the independent variable is the Yelp Dollar-Rating. Location and food category fixed effected are marked below each column. Standard errors are in parentheses and are clustered by category interacted with city or zip-code, as indicated below each column.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A2: Robustness of Main Results

	Linear Specifications			Non-Linear Specifications		
	(1) Ratings	(2) Ratings	(3) Ratings	(4) Ratings	(5) P(Great)	(6) P(Not Bad)
Price (pct.)	-0.158**	-0.044*	-0.154***	-1.000**	-0.588*	-1.922**
	(0.069)	(0.023)	(0.046)	(0.461)	(0.325)	(0.884)
Observations	9333	46941	22513	22512	22512	22512
(Pseudo) Adj. R^2	0.261	0.432	0.058	0.447	0.183	0.474
# of Items	3338	22246	6953	6953	6953	6953
Window	5	15	10	10	10	10

Note: This table reports regression coefficients from six separate regressions. An observation is a (rated) transaction item. The sample only includes reviews given within a narrow window of days around the price change, as indicated below each column. Column (4) is an ordered logit specification, and columns (5) and (6) use logistic regression. Outcomes are indicated in the column headers and described further in the text. 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 A3: The Impact of Prices on Ratings Using Only Multiple Price Changes

	(1) Ratings	(2) Ratings	(3) Ratings	(4) Ratings
Price (pct.)	-0.246**	-0.281**	-0.635***	-0.217***
	(0.125)	(0.119)	(0.239)	(0.075)
Constant	3.328***	3.400***	4.154***	3.284***
	(0.247)	(0.231)	(0.484)	(0.153)
Observations	3126	3748	1214	8319
Adjusted R^2	0.240	0.256	0.246	0.228
# of Items	1047	1021	296	2323
Cutoff	5	10	20	10

Note: This table reports regression coefficients from six separate regressions. An observation is a (rated) transaction item. The sample only includes reviews given around multiple price changes within restaurant, the minimal numbers of weekly price changes are indicated below each column. Outcome is transaction 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 A4: The Impact of Prices on Ratings Using Cross-partner Variation

	(1) Ratings	(2) P(Great)	(3) P(Not Bad)
Price (pct.)	-0.075**	-0.044	-0.031*
	(0.038)	(0.028)	(0.016)
Constant	2.879***	0.862***	1.017***
	(0.070)	(0.053)	(0.029)
Observations	14795	14795	14795
Adjusted R^2	0.060	0.044	0.070
# of Items	4162	4162	4162

Note: This table reports regression coefficients from six separate regressions. An observation is a (rated) transaction item. The sample only includes reviews given around price changes in which there is a lag between different d Outcome is transaction 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 the lagged price change. Standard errors are in parentheses and are clustered at the item-group level.
* significant at 10%; ** significant at 5%; *** significant at 1%

B.2 Data and Data Cleaning

Data for motivating evidence This dataset includes a snapshot of Yelp Star Ratings for all restaurants in Los Angeles, New York, and Houston, as of January 2019. While the Yelp Star Ratings presented to users are rounded to the nearest half star, I use the underlying, continuous, rating. Data includes restaurant city, zip-code, and food category. There are over 244 unique food categories in the data, and some, such as Japanese food and sushi, are clearly not mutually exclusive and are likely to be decent substitutes. In order to control for those similarities, in an additional specification we aggregate food categories to 10 main categories.[1]

Data for main analysis We use YTP data on food orders from 2013 through January 2019. We restrict our attention to orders conducted in the US. We exclude orders that were not completed or canceled by the user, as well as orders from businesses that were marked as fake or fraudulent. Price data excludes taxes or delivery fees. We do not include tips, discounted items, or orders in which a coupon was applied. We also exclude items where the description suggests item modification or additional fees.

To identify price changes, we develop a simple algorithm for data cleaning and sample selection. First, we exclude items for which we have less than ten observations or have no price variation. Second, we omit prices that appear five or less times in our data. This step

[1]The ten categories are: Asian, pizza, Mexican & Latin, European, Arab-Indian, meat & seafood, American, coffee & pastry, sandwiches, and other.

helps clean out noise in the data. Third, we define the price-time-period as the interquartile range (date-wise) plus half the interquartile range, i.e. if we observe an item-price with 25% of observations prior to February 10 until and 25% following February 20, then we define the price-time-period as February 5 to 25 (10-(20-10)/2, 20+(20-10)/2). We then exclude all price points that fall outside the price-time-period. This step pins down the period in which the price was prevalent. Fourth, we allow for transition periods (lags) in which there are two prevailing prices, but: (1) only allow for overlap for up to 10 days, (2) one price (usually the lower one) must prevail in the period prior to the overlap and the other price prevail in the period following the overlap, and (3) most price observations for both price points occur outside of the overlap. In cases where two or more prices prevail, we omit all prices from the analysis.

As a robustness test, we alter the last step to exclude only the higher price of the two. The rationale behind this slightly more lenient definition is that item modification usually results in a price increase, and thus the lower price is more likely to be the baseline price. Finally, we test an alternative sample selection algorithm in which we completely abandon the assumptions above: We do not mark any observations as "noise" and instead drop all items with more than two price, we then drop all items where there is any overlap, even in transition, between the two prices. This specification is far more restrictive, as it does not allow for price variation outside of the coded price change.

B.3 Additional Robustness Tests and Specifications

This appendix discusses the robustness checks conducted for the main results. All relevant figures and tables are in Appendix B.1.

Table A1 presents the formal analysis of the motivating evidence. We regress Yelp Star Ratings on the Dollar Sign Ratings. Coefficients should be interpreted as the impact of the relevant Dollar Rating on Star Rating compared to one-dollar rating (the omitted category). For instance, in Column (1), two-dollar is correlated with a *decrease* of -0.025 stars and this effect is only marginally statistically significant. three- and four-dollars are correlated with increases of 0.078 and 0.139 stars above the one-dollar rating, respectively. Back of the envelope calculation suggests that those imply increases of about 2.2% and 3.5% for the median business.

The different columns control for different combinations of geographic location (city or zip codes), and food categories, as described in Appendix B.2. The qualitative results are robust across columns: Two-dollar restaurants are generally not higher ranked compared to one-dollar restaurants. In fact, the impact of additional dollar sign seems, if anything, to have a weak negative effect (Columns (1), (2) and (5)).

Moving to a three- or four-dollar restaurant has a more positive effect on ratings, with four-dollar restaurants rated marginally better than three-dollar ones. Note, however that only 1% of all restaurants receive a four-dollar rating, and hence these two levels are aggregated in Figure 2. The average magnitude across columns is approximately 0.14, which is about 3.5% of the median Star Rating.

Figure A3 presents the motivation for the cleaning algorithm. Panel A presents items with "good" price variation; in each period there is only one prevailing price and a clear-cut transition. This is the ideal data for our purposes. Unfortunately, much of the data we have looks more like Panel B: It is clear that there are some dominant prices in each period and price changes are sharp and clear. We can, however, see that there are additional prices in each period, adding noise to the data. These additional price points appear sporadically, and are, for the most part, above the dominant price. For these items, we discard observations at non-dominant prices. Finally, Panel C presents price variation which doesn't have clear price changes. We discard all of these items.

Table A2 presents the robustness tests for the main results. In Columns (1) and (2) we change the size of the window around price changes. Recall that in the main results presented in Table 2, we include observation within 10 days of the price changes. In contrast, Columns (1) and (2) include feedback received for orders which took place 5 and 15 days from the price change, respectively. As we can see in Column (1) when narrowing the window to 5 days the estimated effect of log price is -0.158, slightly larger in magnitude than the main specification. In Column (2), we broaden the window to 15 days. the estimate effect is attenuated, -0.044, and is marginally statistically significant (P-value is 0.056). This result is consistent with the motivating evidence present in the text; without controlling sufficient control for time-varying item-level changes, the estimated relation between prices and ratings becomes null.

In Column (3) we exclude the transaction-level control, X_{jt} in Equation 2.2. We are particularly interested in excluding the indicator for whether an order arrived late. The reason is that this indicator is based on consumer feedback and it could be the case that consumer reporting is, at least partially, affected by other factors (such as the rating they intend to give or even price), which would bias our results. Luckily, the estimated coefficient is actually larger than the main specification and is significant at 1%. Note, however, that the adjusted R^2 is substantially reduced when excluding these controls, as expected.

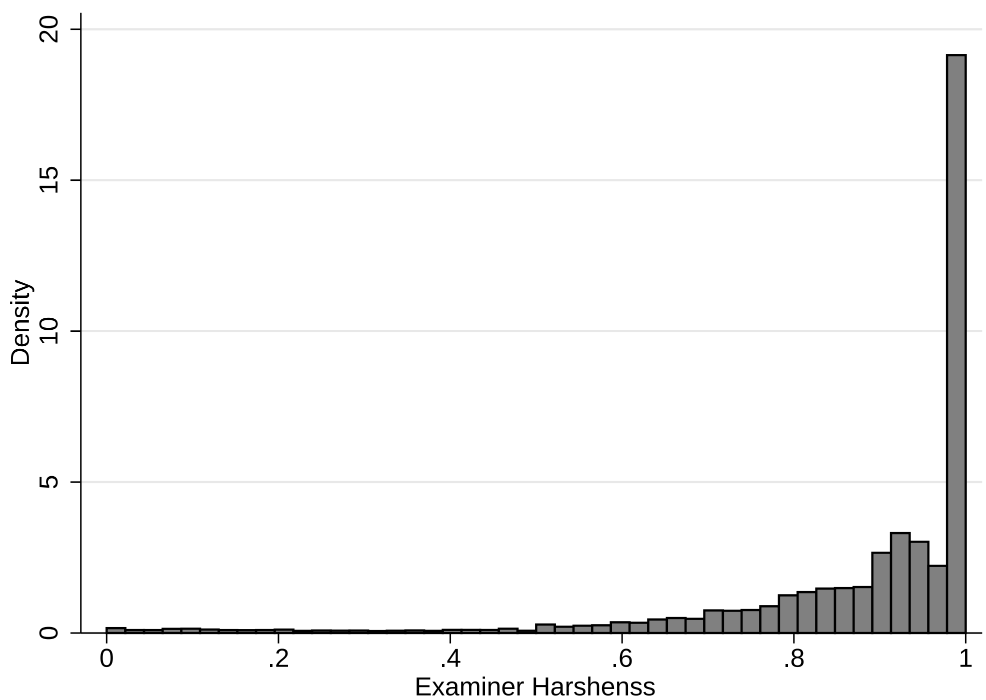
Columns (4)-(6) are equivalent to Columns (1) and (4)-(5) in the Table 2, with non-linear specifications. In particular, ratings for orders are given on an ordinal scale: "Bad", "Good", and "Great", which we code as 1, 2, and 3, respectively. However, it is not clear that "Good" is worth twice as much as "Bad" and two-thirds of "Great". Columns (4)-(6) address this issue by using non-linear specifications. We run an ordered logit regression in Column (4) and conditional logit regressions in Columns (4)-(5). The impact of price on ratings remain qualitatively similar and statistically significant.

Appendix C

Persistence and the Gender Innovation Gap: Evidence from the U.S. Patent and Trademark Office

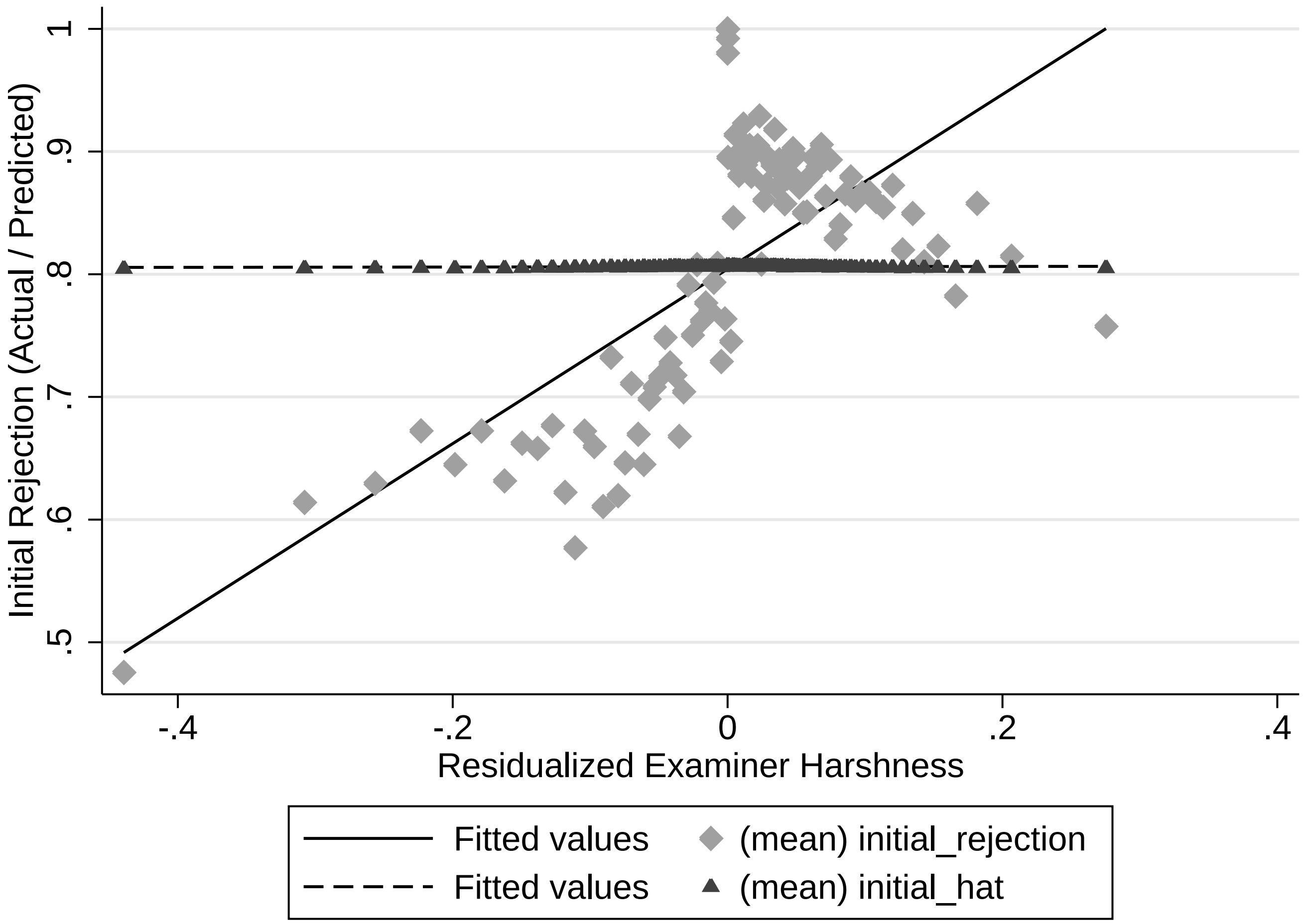
C.1 Appendix Figures and Tables

Figure A1: Distribution of Examiner Harshness by Initial Rejection



Note: This figure shows the raw distribution of patent initial rejection rates.

Figure A2: Examiner Harshness and (Predicted) Initial Rejection (Residualized)



Note: This figure relates our examiner leniency measure, residualized by Art Unit-by-application year fixed effects, to two variables: (1) the initial rejection rate and (2) the *predicted* rejection rate, where we predict a rejection as a function of predetermined observable characteristics: the number of innovator applicants, the proportion of female innovators on an application, whether an application is assigned to an employer, etc.

Table A1: Robustness of Main Results

	(1)	(2)	(3)
	Panel A: Effect on Next Step		
Female X Initial Rejection	-0.015***	-0.098***	-0.043***
	(0.004)	(0.006)	(0.004)
Dependent Var. Mean	0.70	0.71	0.70
	Panel B: Effect on Patent Granted		
Female X Initial Rejection	0.002	-0.068***	-0.023***
	(0.002)	(0.004)	(0.003)
Observations	1031848	915315	1029535
# of Clusters	38682	38682	38682
Dependent Var. Mean	0.88	0.88	0.88
Female Definition	Female Dummy	Single Gender	Half Female
Fixed Effects	Art Unit	Art Unit	Subclass

Note: This table reports regression coefficients from six separate regressions. An observation is a patent application. Outcomes are denoted at sub-headers. We instrument for initial rejection using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. All regressions include lower order interactions. Fixed effects are indicated below each column. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A2: Robustness- Effect of Initial Rejection on Patent Granted & Next Step (Final as IV)

	(1)	(2)	(3)	(4)
	Panel A: Effect on Next Step			
Female X Initial Rejection	-0.050***	-0.053***	-0.044***	-0.091***
	(0.008)	(0.010)	(0.007)	(0.009)
Dependent Var. Mean	0.88	0.86	0.88	0.88
	Panel B: Effect on Patent Granted			
Female X Initial Rejection	-0.104***	-0.086***	-0.079***	-0.100***
	(0.018)	(0.019)	(0.016)	(0.019)
Observations	1031848	478987	1031848	1031848
# of Clusters	38682	38522	38682	38682
Dependent Var. Mean	0.70	0.69	0.70	0.70
Female Definition	Proportion	Solo	Half Female	All Female

Note: This table reports regression coefficients from eight separate regressions. An observation is a patent application. Outcomes are denoted in sub-headers. We instrument for initial rejection using examiners' leave-out mean patent denied rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A3: Mechanism- Persistence over Rounds

	(1)	(2)	(3)
Female X Rejection	-0.290***	-0.032***	-0.042***
	(0.016)	(0.003)	(0.004)
Round X Female X Rejection	0.235***		
	(0.016)		
Female X Rejection X Round 2		0.077***	0.091**
		(0.023)	(0.036)
Female X Rejection X Round 3		0.014	0.092
		(0.042)	(0.076)
Female X Rejection X Round 4		0.168*	0.227
		(0.100)	(0.148)
Female X Rejection X Round 5		-0.051	0.044
		(0.145)	(0.353)
Observations	1865280	1865280	816983
# of Clusters	38682	38682	38529
Dependent Var. Mean	0.84	0.84	0.83
Female Definition	Half Female	Half Female	Solo

Note: This table reports regression coefficients from six separate regressions. An observation is a patent application. Outcomes is a dummy variable for whether a patent was granted. We instrument for rejections using examiners' leave-out mean initial rejection rate for all other applications within art unit-year. Definitions of the Female variable are denoted below each column and are described in the text. Round is the number of initial application plus number of appeals /amendments. All regressions include art unit-year fixed effect and lower order interactions. Standard errors are in parentheses and are clustered at the examiner-year level.
* significant at 10%; ** significant at 5%; *** significant at 1%