

Do Employers Learn from Public, Subjective, Performance Reviews?

Alex Wood-Doughty*

University of California, Santa Barbara

October 2016

Abstract

Much of the new “gig economy” relies on reputation systems to reduce problems of asymmetric information. In most cases, these reputation systems function well by soliciting unbiased feedback from buyers and sellers. However, certain features of online labor markets create incentives for employers to misreport worker performance. This paper tests whether employers learn about worker productivity from public, subjective, performance reviews using data from a large online labor market. Starting with a simple model of employer learning in the presence of potentially biased reviews, I derive testable hypotheses about the relationship between public information and wages, worker attrition, and contract renewals. I find that these public reviews provide substantial information to the market and that other firms use them to learn about the productivity of workers. I also find evidence that these reviews affect how long workers stay in the labor market. Finally, using data on applications, I provide evidence of a mechanism for informative reviews. I show that workers punish firms that leave negative reviews by refusing to work for them again. Together, this body of evidence suggests that reputation systems in online labor markets provide significant information to both workers and firms and help reduce problems of asymmetric information.

JEL Codes: D82 D83 J31 J49

*I thank Peter Kuhn for invaluable help with this project. I also thank Doug Steigerwald, Dick Startz, and seminar participants at the ZEW Workshop on Atypical Employment, the Trans-Pacific Labor Seminar, the WEAI Annual Conference, Berkeley, and UCSB for helpful comments and suggestions. I gratefully acknowledge financial support from the NET Institute (www.NETinst.org).

1 Introduction

The past 15 years have seen the rise of the “gig economy,” with “tens of millions of Americans involved in some form of freelancing, contracting, temping or outsourcing” (Scheiber, 2015). Katz and Krueger (2016) find “that the percentage of workers [in the U.S.] engaged in alternative work arrangements—defined as temporary help agency workers, on-call workers, contract workers, and independent contractors or freelancers—rose from 10.1 percent in February 2005 to 15.8 percent in late 2015”. This increase in freelancers and independent contractors has been facilitated by the development of technological platforms that can bring together individuals (e.g. eBay, Uber, and Airbnb) and allows for more flexible working arrangements for both workers and firms. These platforms allow people on opposite sides of the world to transact and have the potential to greatly increase welfare by reducing transaction costs.

One of the main requirements for the success of the new platforms is reputation: “successful online marketplaces have scaled because they have created well-designed reputation systems that allow users to identify trusted community members to interact with” (Stewart, 2014). Without a reputation system in place, it would be absurd to send money to someone on eBay or get in the car of an Uber driver. This paper tests the functionality of the reputation system of an online labor market. Similar to other online marketplaces, the buyer (an employer) leaves feedback about the seller (the worker). This feedback allows other buyers to learn about the quality of the seller. However, employers in online labor markets have an incentive to bias their feedback. Because rehiring is frequent and an individual review can have a large impact on a worker’s reputation, employers can gain by leaving a negative review and then rehiring the worker at a lower price. If this is the case, it is not obvious that feedback in online labor markets provides useful information to other buyers, and thus, that the reputation system is functional.

I start by presenting a simple theoretical framework of employer learning that allows for uninformative feedback. This model generates testable hypotheses about the relationship between wages and reviews. Using data from a large online labor market, I find strong evidence that public performance reviews do provide information to the market. I find that outside firms use reviews to learn about worker productivity and that this learning happens quite quickly.¹ These reviews also affect how long workers stay in the labor market. Together, this body of evidence suggests that, despite an incentive to bias reviews, employers are leaving public performance reviews that do provide information about workers’ productivity. I test two possible mechanisms for this result: workers refuse to return to employers who have left them poor reviews,

¹I will use employer and firm synonymously throughout.

and firms are concerned about their wider reputation. Using data on applications, I show that workers are unwilling to renew employment contracts with firms that left them a negative review. This will cause firms with a desire to rehire a worker to leave an informative review. I also find that the wage a firm has to pay to hire a worker depends on the previous reviews that the firm has left, implying that firms' reputation is important. Despite these mechanisms, there is still an incentive for firms to behave strategically. I investigate one possible source of this strategic behavior by looking at review comments: firms may be able to leave a negative review score with a positive comment to convince workers to work again. However, I find no evidence of any strategic behavior with regard to comments. I conclude that the reputation system in this online labor market is quite functional and that both rehiring considerations and firm reputation help overcome the incentive for employers to bias their reviews.

This paper contributes to the growing literature on reputation systems.² One of the first empirical studies of a reputation system comes from Resnick and Zeckhauser (2002), who document some empirical facts about eBay using cross-sectional data. More recent studies of eBay have achieved cleaner identification using field experiments (Resnick et al., 2006), panel data (Cabral and Hortaçsu, 2010), and policy changes (Hui et al., 2015). All of these studies find evidence that the reputation system on eBay is functioning. Nosko and Tadelis (2015) and Dellarocas and Wood (2007) document problems with eBay's reputation system, however, these problems do not completely invalidate the system.

There has also been some recent research into other platforms. Pallais (2013) is the closest paper to this one. She runs a field experiment in an online labor market and finds that providing jobs to inexperienced workers significantly improves their later outcomes. Her focus is on the effect of experience on future outcomes and is not directly testing if reviews are informative. Benson et al. (2015) run a field experiment in a different online labor market showing that the employer's reputation affects their ability to hire workers. Fradkin et al. (2016) show that there is evidence of some biased reviews on Airbnb, but as a whole their reputation system works well. Luca and Zervas (2015) and Mayzlin, Dover and Chevalier (2014) document the case of fake reviews on Yelp and Tripadvisor, respectively. While the existence of fake reviews is obviously bad for reputation systems, the fact that companies are posting them suggests they believe that their reputation matters.

Finally, there is an extensive marketing literature on "word-of-mouth" in terms of online product reviews (Chevalier and Mayzlin, 2006; Chen and Xie, 2008; Zhu and Zhang, 2010). Overall, there is a significant body of evidence that reputation systems do a good job of reducing moral hazard and adverse selection. However, it is not

²See Tadelis (2016) and Luca (2016) for recent surveys.

obvious that the reputation systems in online labor markets should follow this trend, as there is an incentive for buyers to bias their feedback.

I propose four characteristics of a market that determine whether there is an incentive for buyers to bias feedback. First, the price of the good must be a function of the feedback. On Uber and Lyft, leaving a negative review of the driver will only impact whether that driver is allowed to continue, as the price of a ride is set centrally by the company. Second, the buyer must want to make additional purchases from the seller. There is no reason to leave biased feedback of an apartment on Airbnb if you are never planning on staying there again. Third, there must be limited other information about the seller. There is not a strong incentive to bias feedback for most goods on Amazon.com, since there is so much additional information about the product and the seller. This also suggests that there is limited incentive to bias feedback of workers in online labor markets with lots of previous reviews. Finally, the feedback must be public information. In the employer learning literature, there is evidence that within-firm performance reviews provide information to the firm (Kahn and Lange, 2014). However, these reviews are private to the firm so there is no incentive to leave dishonest reviews. Outside firms do not observe the feedback, so they cannot bid away workers with good reviews. Table 1 summarizes these examples with predictions.³

Table 1: Characterization of Markets

	Uber	Airbnb	Amazon	Within-Firm Review	Online Labor Market
Price is a Function of Review	<i>X</i>	✓	✓	✓	✓
Repeat Purchases	✓	<i>X</i>	✓	✓	✓
Limited Information	✓	✓	<i>X</i>	✓	✓
Information is Public	✓	✓	✓	<i>X</i>	✓
Incentive to Bias?	No	No	No	No	Yes

Online labor markets are distinct from many other markets in that they exhibit all four characteristics. Wages are a function of the feedback the worker has received. Employers potentially want to rehire the same worker. There is limited information about the productivity of the worker, and all of the feedback is public information. The combination of all four factors implies that employers have an incentive not to provide accurate feedback. If you just hired someone who was excellent at their job, why would you provide your competitors with this information? The argument stems from the work of Milgrom and Oster (1987) who find that firms have an incentive to

³There might be specific cases where there does exist an incentive to bias in other markets (e.g. renting the same apartment on Airbnb or reviewing a third-party seller on Amazon with limited feedback). However, these are special cases and there is always the incentive to bias in an online labor market.

hide their talented workers by not providing public information about their skill (i.e. promoting them).

This paper also contributes to the literature on employer learning. This literature dates back to the work of Farber and Gibbons (1996) and Altonji and Pierret (2001) who propose that firms observe a noisy, publicly available signal of worker productivity. However, we might believe that the employing firm has more and/or better information about a worker than outside firms. There is also evidence that outside firms learn about the productivity of workers. There are a number of different mechanisms proposed for this learning. Waldman (1984) proposes that outside firms learn about productivity by observing promotions within another firm. Pinkston (2009) argues that private information is reflected in a worker's wage and this information is passed to other employers when the worker makes a job transition. Finally, Kahn (2013) shows that outside firms are able to learn more about workers whose jobs require more outside communication. In this paper, I study a more direct mechanism for outside firms to learn about the productivity of a worker: public performance reviews. This is the first paper to study employer learning from this new type of information, which will become more important as the labor market continues to become digitalized.

To motivate the role of employer learning in this market, I present evidence for the three main results from the symmetric employer learning literature. The first is that the cross-sectional variance of wages should increase with experience. When a worker is first hired, a firm does not know much about his productivity, so, conditional on observables, all workers are paid similarly. Over time, firms observe the workers performance and their wage will reflect their productivity. Highly productive workers will have more wage growth than lower productivity workers, thus the variance of overall wages will increase over time. Figure 1.1 plots the cross-sectional variance of wages for the first 20 jobs of a worker's career.⁴ There are clear increases in the variance of wages over time ($p < 0.01$ for the slope coefficient).

The other two results describe the relationship between wages and worker characteristics. The correlation between wages and easily observable characteristics should decline over time, while the correlation between wages and more accurate measures of productivity should increase. Without much information about the worker, wages are set based on easy to observe characteristics. However, these characteristics do not perfectly capture worker productivity. As firms learn about a worker's productivity through work experience, if that learning is reflected in the reviews, we would expect the effect of reviews to increase over time. Figure 1.2 plots the marginal effect of an easily observable characteristic (whether the worker is from a high income coun-

⁴Only workers who stay for at least 20 jobs are included.



Figure 1.1: Variance of Wages

try) and the effect of the average review.⁵ The importance of the worker’s country in determining wages decreases over time, while the effect of the average review score increases.



Figure 1.2: Marginal Effects of Worker Characteristics on Wages

These results suggest that the public performance reviews are informative and the labor market learns about the ability of workers over time. However, this does not show that firms are learning directly from the reviews. The previous results are consistent with a model where worker output is publicly observable, or with a model where output

⁵Appendix A.1 provides the details of the specification, which follows Altonji and Pierret (2001).

is private but workers always work for the same employer. To directly test for learning from public performance reviews, I need to explicitly measure the effect of a single review on how much workers are paid by other employers.

The rest of the paper is organized as follows. Section 2 provides some background on online labor markets and describes the data. Section 3 presents a theoretical framework for employer learning and generates testable hypotheses. Section 4 tests these hypotheses. Section 5 investigates the effect of reviews on worker attrition. Section 6 discusses and tests possible mechanisms for these results and Section 7 investigates other possible strategic behavior by firms.

2 Data

2.1 Online Labor Markets

Online labor markets are platforms on which workers are matched to short-term tasks and where workers' output is delivered to employers electronically. The combination of these two features distinguishes this new form of labor market institution from job boards and social networking sites (which only perform the former function, and which focus on formal employment) and from telework, which only does the latter. oDesk.com (which has since rebranded as Upwork.com) had 9.7 million workers and 3.8 million employers, with workers earning almost \$1 billion in 2014 (Elance-oDesk, 2014). Other significant online labor markets include guru.com, freelancer.com, and Amazon's MTurk.com. Online labor markets provide opportunities for marginally attached workers, workers from third-world countries, and workers with flexible hours requirements. A survey of U.S. workers found that 25% of freelancers had a traditional full-time job but were supplementing their income with additional work. Another 26% of freelancers were classified as diversified workers, i.e. they were working part-time at a traditional job and working as a freelancer (upwork.com, 2015). A survey of workers on Amazon's MTurk.com found that 34% of workers were from India and for these workers the money earned was likely to be a primary source of income (Ipeirotis, 2010). Online labor markets make it easy for workers to find employers who are willing to compensate them for a variety of short-term tasks.

This study will use data from oDesk. Workers on oDesk create a profile where they can include relevant information about themselves, including education, outside work experience, and location. They can also take skill tests on oDesk to signal their proficiency at different tasks. The final aspect of their profile comes from performing jobs on oDesk. Employers on oDesk can post jobs. Workers apply to those openings and propose a wage. The employer and worker can then bargain over the wage and

when they agree, they enter into a contract. Every time a worker is hired through oDesk, the job information and wage are posted on their profile. When the job is complete, the employer has the option of leaving a review, which is also posted on the worker's profile.⁶ The employer grades the worker out of five stars in six different categories: Availability, Communication, Cooperation, Deadlines, Quality and Skills, with the average of the six scores being shown as the overall score, which I will call the review score. Perhaps surprisingly, reviews are left around 75% of the time. oDesk facilitates matches by allowing workers and firms to search for each other with very detailed filters.

2.2 Data Description

The data consists of the universe of oDesk workers who were active in the administrative job category between January 1, 2015 and April 25, 2015. I observe every job these 15,684 workers have done on oDesk from when they first joined through April 25, 2015, with the review they receive. I also observe their profile, which includes their country, education, oDesk test scores and previous experience. From this data, I construct a panel where each observation is a completed job by a worker, henceforth referred to as a job. I limit my sample to the first 20 jobs of each worker's career. Table 2 provides summary statistics for the workers. The majority of workers are from Lower Middle Income countries, with India and the Philippines accounting for almost 60% of all workers.⁷ Nearly half of all workers report having at least a Bachelor's Degree. I focus my analysis on a single category: administrative jobs. Administrative jobs consist of data entry, web research, personal assistant jobs. These jobs are generally low-skilled and pay a relatively low hourly wage.⁸ I choose to look at administrative jobs because they are more homogenous than other categories and the majority of them are hourly jobs.⁹ The average (partial) career is seven jobs with each job being a fairly significant time commitment: almost a month long and 40 hours per week. This differentiates this online labor market from Amazon's Mechanical Turk, where the jobs are very short term (a few seconds to a few minutes). Here the employers develop a significant relationship with their workers, and can learn about their productivity. Finally, workers rarely have more than one just at a time, so their career on oDesk is fairly sequential.

Figure 2.1 shows the density of review scores across jobs. It is highly skewed towards

⁶Workers also have an option of reviewing their employer. These reviews are simultaneous and blind, so we would not expect there to be a threat of retaliation.

⁷I classify countries according to the World Bank Country Income Classification (World Bank, 2013).

⁸However, the hourly wage is comparable to the average hourly wage in India and the Philippines.

⁹There are two types of jobs on oDesk: hourly and fixed price. I limit the bulk of my analysis to hourly jobs, since I do not observe the time spent on fixed price jobs, and thus, cannot compare them.

Table 2: Summary Statistics for Workers

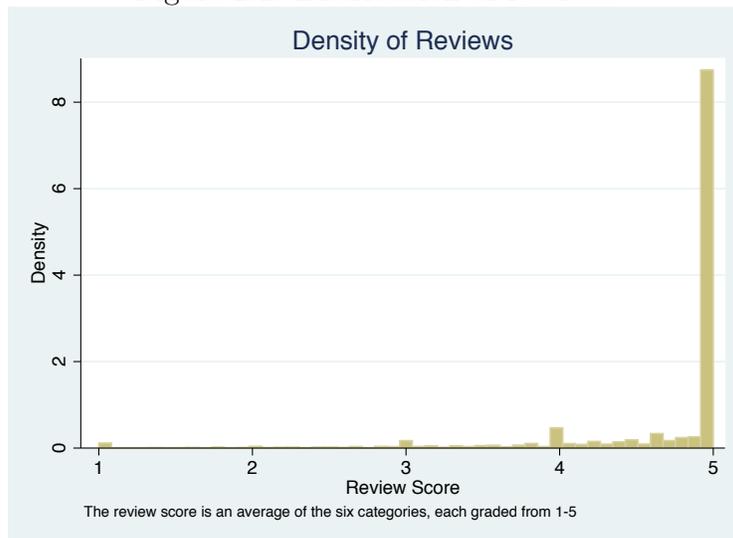
High Income Country: OECD	0.135
High Income Country: Non-OECD	0.00976
Upper Middle Income Country	0.0550
Lower Middle Income Country	0.655
Low Income Country	0.145
Bachelor's Degree	0.498
Master's Degree	0.0752
Number of Outside Experiences	2.670
Number of oDesk Exams	2.136
Average Wage (\$)	5.563
Average Duration (Hours)	172.3
Average Duration (Days)	29.94
Average Career Length	7.258
Average Num Jobs at One Time Conditional on Having a Job	1.243
<i>N</i>	15684

a perfect score. For the majority of my analysis, I will classify the review into four types: Good, Bad, Really Bad, and No Review according to:

$$\text{Review Type} = \begin{cases} \text{Good} & \text{if Score} = 5 \\ \text{Bad} & \text{if } 4 \leq \text{Score} < 5 \\ \text{Really Bad} & \text{if Score} < 4 \\ \text{No Review} & \text{if no review was left} \end{cases} \quad (1)$$

While it may seem strange to classify a score of 4.85 out of 5 as a bad review, it is only the 29th percentile of the distribution. Similarly, a score of 3.95 is the 10th percentile of the review distribution, so these really are “bad” and “really bad” reviews.

Figure 2.1: Distribution of Reviews



3 Theoretical Framework

To formalize the relationship between wages and reviews, I develop a simple model of an online labor market. Workers have a publicly observable, fixed characteristic, x_i , and their productivity is a linear function of this observable characteristic and an unobservable, idiosyncratic component: $\theta_i = \gamma x_i + \nu_i$.¹⁰ When a worker is hired by a firm, that firm observes a noisy signal of productivity: $y_{it} = \theta_i + \epsilon_{it}$. After the completion of the job, the firm leaves a review, r_{it} . Consider two possible types of review-setting. Define an informative review as any review that is a function of the

¹⁰I am assuming no human capital accumulation for simplicity, although it does not fundamentally change the predictions.

observed signal of productivity: $r_{it} = g(y_{it})$ for some weakly monotonically increasing function g . Thus, honest revelation ($r_{it} = y_{it}$), shading by a constant, known factor ($r_{it} = 0.7 * y_{it}$), and a threshold model ($r_{it} = \begin{cases} 5 & \text{if } y_{it} \geq \bar{y} \\ 1 & \text{if } y_{it} < \bar{y} \end{cases}$) are all informative reviews. Define an uninformative review as any review which contains no information about the observed signal of productivity: $r_{it} \perp\!\!\!\perp y_{it}$. Examples of uninformative reviews include choosing a review at random ($r_{it} \sim U[1, 5]$) and always leaving the same review ($r_{it} = 3$). In this competitive marketplace, wages are a function of the expected productivity of the worker, conditional on the information available: $w_{ijt+1} = \mathbb{E}[\theta_i | I_{it}] + \eta_{ijt+1}$ where η_{ijt+1} is some job specific component that might depend on firm j and I_{it} is the information about the worker that is available to the firm. Finally, workers are assumed to stay in the market for their entire career.

Consider two possible worlds: one where all reviews are uninformative and one where at least some reviews are informative. An uninformative world is one in which reviews contain no information about the worker and therefore have no effect of wages:

$$w_{ijt+1} = \mathbb{E}[\theta_i | x_i] + \eta_{ijt+1} \quad (2)$$

This is the extreme case where no one believes the reviews left by other firms.¹¹

In contrast, in an informative world, at least some of the reviews contain information about the worker, so they affect wages:

$$w_{it+1} = \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it}] + \eta_{ijt+1} \quad (3)$$

An informative world is one in which at least some employers are providing their private information to the market.¹² Although providing private information is potentially costly to firms, there is a number of reasons we might expect this to happen. Firms might face a “lying cost” Gneezy (2005) of leaving a review that differs from the observed productivity. Firms might also be concerned about reputation effects: other workers may not want to work for a firm who leaves many poor reviews. Finally, a worker may not want to be rehired by an employer that left a poor review. All of these factors potentially enter into the employer’s decision process.

The goal of this theoretical framework is to isolate the effect of the most recent review of wages. To do this, I first difference Equations (2) and (3). This will (mostly)

¹¹This would also be the case if firms did not leave reviews at all.

¹²A more formal treatment of this model is presented in Appendix A.2.

difference out the effect of previous reviews.¹³ In an uninformative world:

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i|x_i] + \eta_{ijt+1} - (\mathbb{E}[\theta_i|x_i] + \eta_{ikt}) \quad (4)$$

where j indexes the firm that hired worker i in period $t + 1$ and k indexes the firm in period t . In an informative world:

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i|x_i, r_{i1}, \dots, r_{it-1}, r_{it}] + \eta_{ijt+1} - (\mathbb{E}[\theta_i|x_i, r_{i1}, \dots, r_{it-1}] + \eta_{ikt}) \quad (5)$$

These two wage equations yield the following hypotheses about the relationship between wage changes and worker characteristics. Without loss of generality, assume that $\gamma > 0$, so that the observable characteristic positively affects productivity. All proofs are in the Appendix.

Proposition 1. *In an informative world, the marginal effect of the most recent review on the conditional mean of the change in the wage is strictly positive and decreasing with t . In an uninformative world, the marginal effect is zero.*

In an informative world, if a worker receives a good review in their previous job, they will receive an increase in their wage. This increase will decline over time, however. The first few reviews are providing lots of new information about the worker's productivity. However, once the worker has a large number of reviews, the marginal effect of an additional review is lowered. In an uninformative world, reviews provide no information about the worker, so they have no effect on wages.

Proposition 2. *In an informative world, the marginal effect of the observable characteristic on the conditional mean of the change in the wage is strictly negative and increasing with t . In an uninformative world, the marginal effect is zero.*

Consider two workers with no reviews and one with more education than the other. The more educated worker will receive a higher wage in the first period. In an informative world, if both workers receive the same review, the worker with the lower education will see a larger change in their wage than the worker with more education. This effect will taper off over time as more and more of the weight is placed on the reviews. I can test these two propositions by looking at data on reviews and wages.

¹³This also removes any concern about the role of endogenous worker effort. Since the next wage is determined before the worker does any work, the worker's choice of an effort level will not effect the relationship between their previous review and the wage change.

4 Learning by Other Firms

To investigate across-firm learning, I limit my sample to the 80% of jobs in which the worker was never previously employed by the hiring firm. If workers choose to leave the market based on their previous reviews (see Section 5), this will likely attenuate the effects of reviews on wage changes. I then regress the change in the (log) wage on worker characteristics, job characteristics and the previous performance review of the worker. A correlation between previous performance reviews and wage changes suggests that these reviews are providing information about the worker to future employers. To test for changes over time, I include an indicator that equals one for all jobs after the fifth and its interactions with review type. This allows me to distinguish the initial effects of performance reviews from their later effects.¹⁴ I estimate the following equation:

$$\begin{aligned} \Delta \log(w_{ijt}) = & \beta \mathbf{ReviewType}_{it} + \gamma \mathbf{After}_{it} + \delta (\mathbf{ReviewType}_{it} * \mathbf{After}_{it}) \\ & + \rho \mathbf{CountryIncome}_i + \tau (\mathbf{CountryIncome}_i * \mathbf{After}_{it}) \\ & + \lambda \mathbf{X}_{it} + \mu \mathbf{J}_{it} + \nu_i + \kappa_j + \epsilon_{ijt} \quad \mathbb{E}[\epsilon_{ijt} | \mathbf{Z}_{ijt}] \end{aligned} \quad (6)$$

where $\mathbf{ReviewType}_{it}$ are indicators for the type of the most recent review and $\mathbf{CountryIncome}_i$ are indicators for the income classification of the worker’s country. \mathbf{X}_{it} are worker characteristics, such as education, exam scores, and previous review scores and \mathbf{J}_{it} are job characteristics, which include indicators for different words that show up in the job titles.¹⁵ ν_i are worker fixed effects and κ_j are firm fixed effects, which are included depending on the specification. Finally, \mathbf{Z}_{ijt} contains all the included variables.

Table 3 shows the results. The results are broadly consistent across specifications. For specification 4, initially there is a significant effect of both really bad and good reviews on the conditional mean of wage changes. On average, receiving a really bad review lowers a worker’s next wage by 10% relative to receiving a bad review. Similarly, receiving a good review increases the worker’s next wage by 4%. There is no significant difference between receiving no review and receiving a bad review. After five reviews, there is no significant penalty for a really bad review nor a bonus for a good review. This provides support for informative reviews from Proposition 1:

¹⁴The results are consistent for a number of different cut-points and non-parametric experience. See Appendix A.5.

¹⁵To generate the job title indicators, I construct a separate dataset of administrative jobs. I construct a document term matrix with the titles of every job. I then run a cross-validated LASSO of the job wage on its document terms (Friedman, Hastie and Tibshirani, 2010). This regularization picks the job title terms that best explain the wage. This procedure will capture differences in wages that are explained by the job title. Given the list of important terms, I create indicators for whether any of the jobs in the main dataset contained those terms.

firms learn very quickly about the productivity of workers, initially wages react to new information about worker productivity, however, after a few periods, new information has a reduced impact. Specification 3 provides limited evidence for Proposition 2. Initially, being from a high income country relative to a low income country decreases your next wage by 11%, but this effect decreases to 9% after five jobs. Similarly, being from a lower middle income country decreases your next wage by 5%, but this effect decreases to 3.5% after five jobs. After only five jobs firms are still willing to pay a premium for workers from higher income countries. While the decreases are not significant at the usual levels, there does seem to be a decline in the amount of the premium. Finally, comparing specifications 2 and 4 suggests there is sorting along a non-wage dimension as well as by wages. Specification 4 is comparing identical workers working for identical firms. Specification 2 does not control for firm characteristics, so workers who receive a good review are willing to take a small pay cut (2%) to work for (presumably) a better firm. Similarly, workers who receive a really bad review can reduce their wage penalty by working for a different type of firm.

5 Worker Attrition

In the previous section, we saw that reviews have an effect on wages early in a worker's career, but this effect is reduced after only a few jobs. It is possible that these public performance reviews also have an effect on worker attrition: whether they continue working in this online labor market. To motivate this idea, I look at the variance of wages over the first 20 jobs for two populations, workers who stayed for 20 jobs and all workers. Figure 5.1 is identical to Figure 1.1. Figure 5.2 shows the variance for all workers, which does not exhibit the same increase over time. This suggests that, while firms are learning about the workers who stay, the workers who stay are different from those who leave: lower productivity workers choose to leave the labor market rather than accept low wages. If this is true, than the results in the previous section have been biased towards zero, as workers who receive poor reviews choose to leave rather than accept negative wage changes, and thus, the true effects of reviews on wage changes are even greater.

Figure 5.3 shows a hypothetical example that is consistent with observed trends in variance. Workers 1 and 2 stay for all three periods. The conditional variance only includes them, and increases over time as their wages diverge. Worker 3 stays for two periods with a constant wage and then leaves, while worker 4 only stays for one period. The unconditional variance includes all four workers and is constant over time. This is because the workers who leave have lower productivity than the workers who stay.

To test the hypothesis that public performance reviews affect worker attrition in

Table 3: Effect of Reviews on Wage Changes

		(1)	(2)	(3)	(4)
Really Bad Review	First 5 Jobs	-0.0291 (0.0177)	-0.0350 (0.0229)	-0.0593* (0.0281)	-0.108** (0.0387)
	After 5 Jobs	-0.00142 (0.00815)	0.00662 (0.00985)	-0.0152 (0.0155)	-0.0320 (0.0197)
No Review	First 5 Jobs	0.0235 (0.0146)	0.0265 (0.0196)	0.0381 (0.0237)	0.0112 (0.0314)
	After 5 Jobs	0.0119 ⁺ (0.00614)	0.0117 (0.00758)	0.0178 (0.0123)	0.0223 (0.0157)
Good Review	First 5 Jobs	0.0202* (0.00958)	0.0192 (0.0127)	0.0507** (0.0155)	0.0404 ⁺ (0.0222)
	After 5 Jobs	0.0134** (0.00514)	0.00950 (0.00630)	0.00712 (0.0103)	0.00205 (0.0130)
High Income Country	First 5 Jobs	-0.0434** (0.0113)		-0.109** (0.0238)	
	After 5 Jobs	-0.0249** (0.00645)		-0.0897** (0.0182)	
Lower Middle Income Country	First 5 Jobs	-0.0278** (0.00896)		-0.0482** (0.0166)	
	After 5 Jobs	-0.0214** (0.00510)		-0.0359** (0.0125)	
Worker Fixed Effects		No	Yes	No	Yes
Firm Fixed Effects		No	No	Yes	Yes
R^2		0.0452	0.130	0.430	0.605
Number of Workers		12924	9939	11036	7434
Number of Firms		43466	42217	12839	11594
Number of Observations		72187	69202	41560	36798

Note: The omitted review type is Bad Review and the omitted country type is Low Income. Regressions contain controls for education, country, previous reviews, previous experiences, exam scores and participation, and job titles. The sample is the first 20 jobs of each worker's career.

⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses



Figure 5.1: Conditional Variance



Figure 5.2: Unconditional Variance

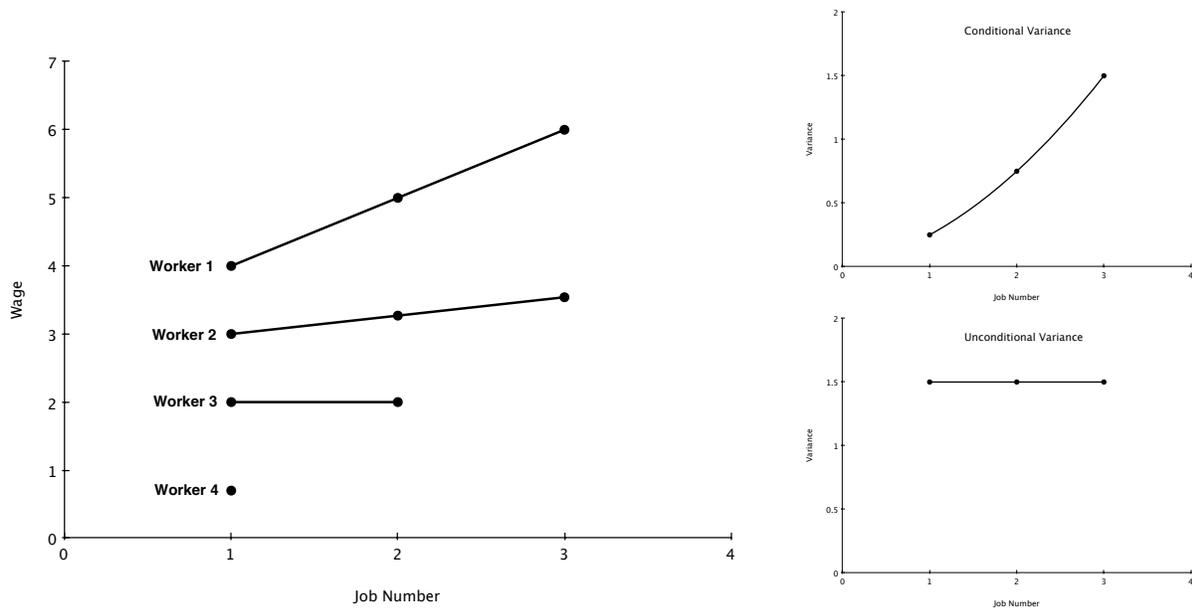
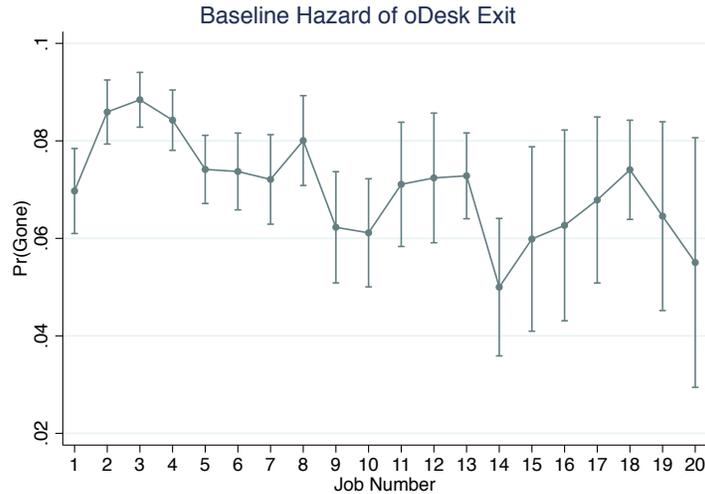


Figure 5.3: Wage Variance Example

an online labor market, I model workers' careers using a discrete-time proportional hazard model. My outcome of interest is whether that worker is hired again on oDesk. This outcome is a function of a fully non-parametric baseline hazard rate as well as worker characteristics. In particular, I allow the overall hazard rate to depend on all the variables from Equation 6, as well as allowing the effects of previous reviews and the worker's average review score to vary over time. Thus, the hazard rate for worker i and job t is given by:

$$\lambda_{it} = \lambda_0(t) * \exp(\beta_t ReviewType_{it} + \gamma_t AvgScore_{it} + \delta \mathbf{X}_{it} + \mu \mathbf{J}_{it}) \quad (7)$$

where both β_t and γ_t vary with t . Figure 5.4 shows the baseline hazard.¹⁶ The probability of leaving the online labor market gradually declines over time.¹⁷ This means that workers are less likely to leave the longer they stay in the market.



Note: 95% confidence bands shown.

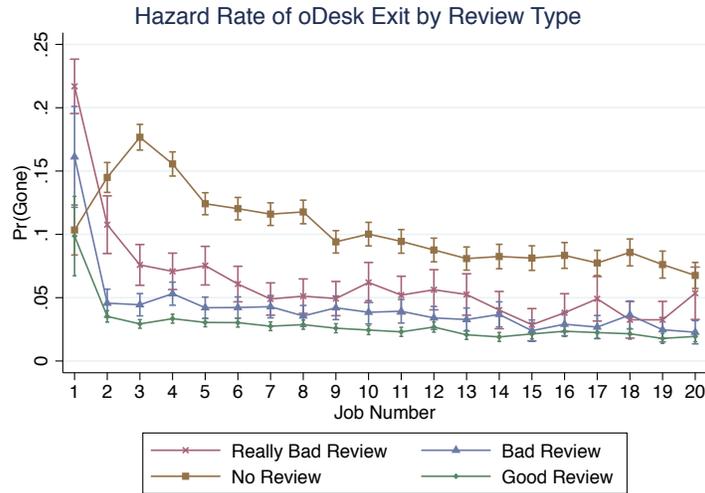
Figure 5.4: Baseline Hazard

To investigate the effect of reviews on attrition, Figure 5.5 plots the hazard rate for receiving each type of review for each job number. All three of the actual review types follow a similar pattern, with a sharp decline in the probability of leaving in the first few reviews. Receiving no review has a much different effect. For the first job, receiving no review is similar to receiving a good review, however, in all subsequent jobs, workers are much more likely to leave the labor market if they do not receive a review. This is consistent with workers having an expectation of receiving some type of

¹⁶I estimate a discrete-time proportional hazard model allowing for normally-distributed unobserved heterogeneity and cannot reject the null hypothesis that there is no unobserved heterogeneity, so my preferred specification assumes no unobserved heterogeneity.

¹⁷ $p < 0.01$ for the slope of the estimated coefficients.

feedback. Since we have already seen the importance of reviews with respect to wages, failing to receive a review may cause workers to become disheartened with the online labor market and leave. Workers who receive a really bad review are over twice as likely to leave as those who receive a good review. After a few jobs, the hazard rates start to converge, suggesting that the type of review does not affect the probability of leaving.



Note: 95% confidence bands shown.

Figure 5.5: Hazard Rate by Review Type

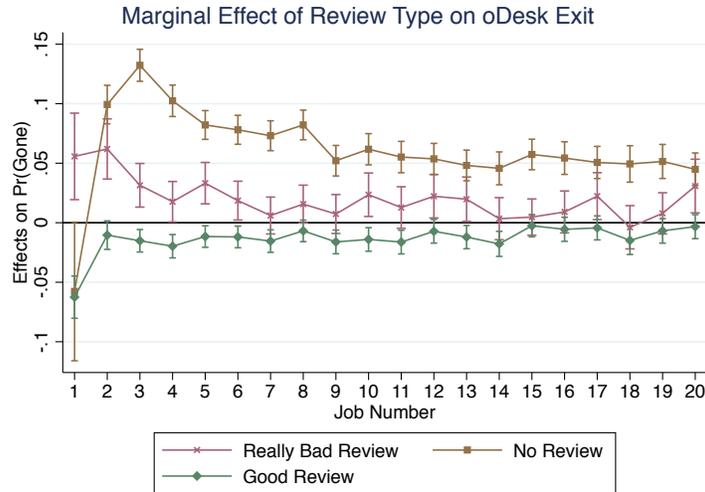
Figure 5.6 plots the marginal effect of the review types relative to a bad review. This plot clearly shows that really bad reviews almost always increase the probability of leaving, while good reviews always decrease the probability. However, these differences decrease over time.

To test this convergence, Figure 5.7 plots the difference between the marginal effects of a good review and a really bad review. While the difference does not go to zero in the first 20 jobs, there is a clear decline in the differences between types of review.¹⁸

Finally, Figure 5.8 plots the marginal effect of the worker’s average review score on their hazard rate. A one point increase in the worker’s average score—going from a 4 to a 5—decreases their probability of leaving by between 2% and 3%. While this is a significant and intuitive result, note that the marginal effect does not change much over time. This result is consistent with firms (and workers) being Bayesian updaters: an individual signal declines in importance over time, however the average of all the signals remains significant.

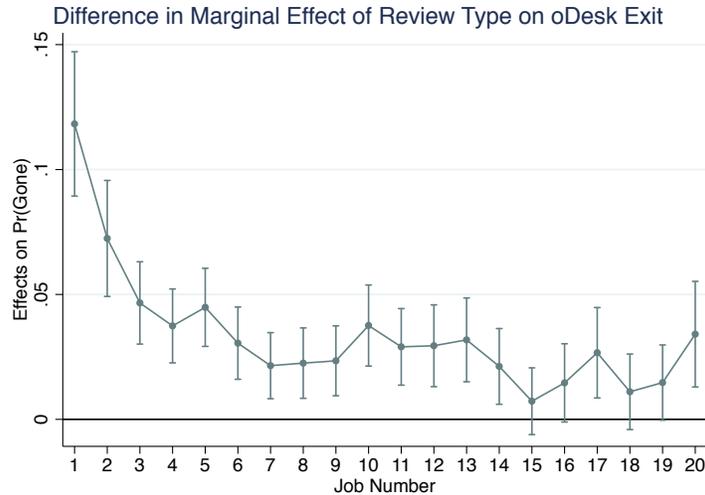
The above results provide additional evidence that performance reviews are provid-

¹⁸ $p < 0.01$ for the slope of the estimated coefficients.



Note: 95% confidence bands shown.

Figure 5.6: Marginal Effects of Review Type

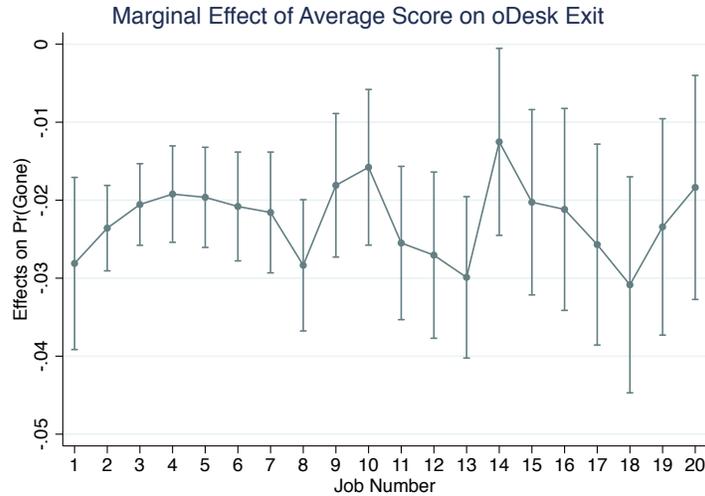


Note: 95% confidence bands shown.

Figure 5.7: Difference in Marginal Effects of Review Type

ing information to the market.¹⁹ If neither firms nor workers believed the reviews, then there should be no effect of a particular review type on the probability of a worker leaving the market. Combined with the results on wages, this suggests that both firms and workers are learning from these reviews and thus, that firms are providing their private information to the labor market. However, it is not obvious why this is the case. I now investigate possible mechanisms that would induce firms to leave informative reviews.

¹⁹Appendix A.6 provides additional evidence that reviews affect the time to find a new job.



Note: 95% confidence bands shown.

Figure 5.8: Marginal Effect of Score

6 Why Do Employers Leave Informative Reviews?

One of the characteristics from Table 1 that was necessary for biased reviews was repeat purchases. In many online marketplaces, such as Amazon, it is straightforward to make a repeat purchase. Labor markets, on the other hand, are two-sided, so both the firm and the worker need to want to match again for there to be a repeat purchase. If workers feel they have been unfairly reviewed, they are unlikely to want to work for that firm again. As one worker put it on the official oDesk forum:

This happens to me sometimes... They hire me, and then give me 4 stars and a review that makes me sound like I should have gotten 6 stars. Then they offer me more work... No thanks, I don't want more 4 star reviews tarnishing my reputation (Daniel C, 2015).

If a worker does not want to work again for a firm, there will be no repeat purchases. As an employer commented:

I think a lot of clients [don't] want to rock the boat, encounter confrontation or jeopardize the relationship moving forward... that's why there are so many five star ratings (Scott E, 2016).

If firms are interested in rehiring the worker, they may be required to leave informative feedback. This suggests a mechanism by which firms are forced to leave informative

feedback.²⁰ To test this mechanism, I turn to application data that provides information on worker decision-making.

6.1 Modeling Contract Renewals

I want to investigate the effect of reviews on subsequent decisions by firms and workers who have previously matched. After a completed job, I observe both worker and firm behavior with regard to subsequent job postings by the same firm. I model the decisions of both workers and firms as a function of their first encounter.

6.1.1 Data

I construct a sample of 11,175 completed administrative jobs. For each job, I observe the worker, firm, and the public performance review. I then look for subsequent job postings (in the next four months) by the firm and applications by the worker to those jobs.²¹ For each application, I see who initiated the application (worker or firm) as well as the outcome of that application. Most applications are worker-initiated, however, the firm has the option of seeking out workers and asking them to apply to their posted job.²² Once there are applications to a job posting, the firm can choose to hire any, or all, of the applicants. There are thus five possible outcomes after a completed job by a worker-firm pair with a subsequent job posting: no application, worker-initiated application and not hired, worker-initiated application and hired, firm-initiated and not hired, and firm-initiated and hired.²³

6.1.2 Model

I model the subsequent decisions of both worker and firm as a function of the previous review. I model this as a bivariate probit, where workers and firms each receive (potentially correlated) shocks and then make their own decision about whether to match. Figure 6.1 shows the decision tree.

If both worker and firm decide to match, I observe an application and a hiring (I am combining the two hiring outcomes into one). If only one wants to match, I see that type of application, but no hire, and if neither wants to match, I do not observe an

²⁰Firms might also be tempted to leave overly positive feedback in the hopes of rehiring a worker. Horton and Golden (2015) investigate the possibility of review inflation on oDesk and find strong evidence that some firms do inflate their reviews.

²¹I limit my sample to only workers who work at least one more job (from any employer) in the future, to control for attrition.

²²While this allows for private offers as in Brown, Falk and Fehr (2004), all job postings are public and third-party enforceable, so it is not obvious what behavior we should expect.

²³If the firm initiates the application and the worker is not hired, this could either mean the worker chose not to apply or that they did apply but were not ultimately chosen by the firm.

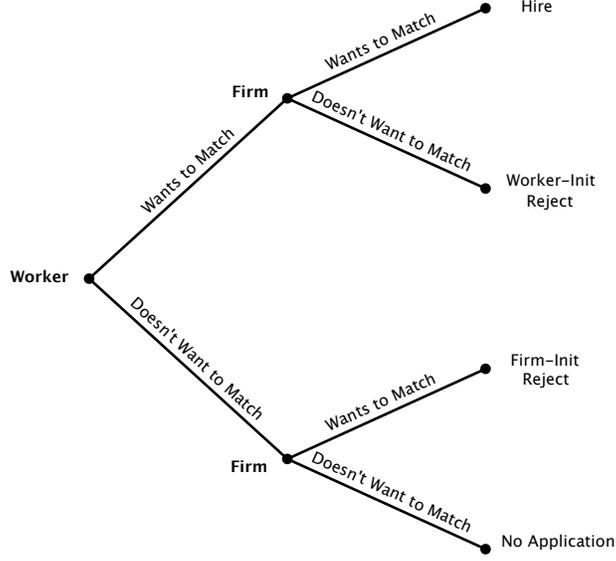


Figure 6.1: Decision Tree for Contract Renewals

application. Both worker and firm decisions depend on the previous public performance review as well as other worker and firm characteristics. The worker's decision is:

$$WantToMatch = \mathbb{1}\{x'\beta + \epsilon_w > 0\} \quad (8)$$

while the firm's decision is:

$$WantToMatch = \mathbb{1}\{z'\gamma + \epsilon_f > 0\} \quad (9)$$

where

$$\begin{pmatrix} \epsilon_w \\ \epsilon_f \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

and x are variables that affect the worker's decision and z are variables that affect the firm's decision. I include the same variables as in Equation 6 for both the workers' and firms' decisions. The likelihood function for the full model is given in Equation 10.

$$f(outcome|\beta, \gamma) = \begin{cases} \Phi(x'\beta)\Phi(z'\gamma) & \text{if Hired} \\ \Phi(x'\beta)(1 - \Phi(z'\gamma)) & \text{if Worker App, Reject} \\ (1 - \Phi(x'\beta))\Phi(z'\gamma) & \text{if Client App, Reject} \\ (1 - \Phi(x'\beta))(1 - \Phi(z'\gamma)) & \text{if No App} \end{cases} \quad (10)$$

where Φ is the standard normal CDF.

6.1.3 Results

Figure 6.2 shows predicted contract renewal rates based on maximum likelihood estimates of Equation 10.²⁴ The effect of no review is similar to a really bad review for both workers and firms and there is a monotonic trend in the effect of review types on subsequent decisions. Both firms and workers are twice as likely to want to match again if the previous review was good than if it was really bad. This suggests that firms are “honest” in their reviews, in the sense that they are more likely to rehire someone to whom they gave a good review. Consistent with the qualitative evidence, workers also select based on the reviews they receive. They are much less likely to apply to a job from someone who has given them a negative review. Finally, the estimated correlation coefficient ($\hat{\rho} = 0.95$) suggests that almost all of the decision-making shock is match specific.

Note that these results are not necessarily identifying the true desires of workers and firms. It is certainly possible that firms would like to shade their reviews (as the theory suggests), but they know that workers will not work be willing to work again if they receive a poor review. However, these results are providing an estimate of the causal link between reviews and subsequent hires. Thus, these estimates provide evidence for a mechanism that induces firms to leave informative reviews. Since the incentive to bias reviews only exists when there is a possibility of a repeat purchase, if workers remove that possibility, then there is unlikely to be biased reviews.

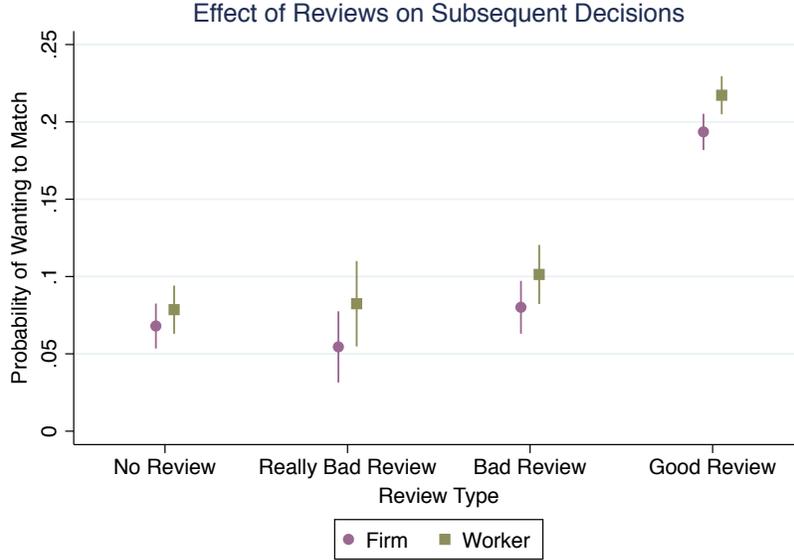
6.2 Effect of Firm Reputation

Another possible mechanism that will prevent firms from biasing their feedback is their overall reputation in the marketplace. A recent study of Amazon’s Mechanical Turk (another online labor market) found that employers with a good reputation were twice as likely to attract new workers as employers with bad reputation (Benson et al., 2015). It is likely that this is important on UpWork as well. As a worker claims:

Whenever I consider applying for a contract I take a long, hard look at the client’s feedback history. It doesn’t worry me when they are new but clients who consistently leave bad feedback (and anything below a 4 is, in my book, very poor) are likely to be either incredibly unlucky, incredibly difficult or incompetent at hiring and running a contract. So I do not bid on their projects and I do not accept offers from them (Petra R, 2013).

To test the effect of firm reputation, I regress (log) wage on a number of firm, worker, and job characteristics. In particular, I am interested in two measures of firm reputation. The first is the average feedback the firm has left for workers (what Petra was

²⁴Table 8 in Appendix A.7 provides the coefficients.



Note: 95% confidence bands shown.

Figure 6.2: Effect of Reviews on Subsequent Decisions

talking about above) as well as the average feedback those workers have left about the firm. Importantly, I include controls for a number of different worker characteristics as well as worker fixed effects, so I can interpret the results as the effect of the firm characteristics on the conditional mean of the wage the firm must pay for identical workers. I estimate the following equation:

$$\begin{aligned}
 \log(w_{ijt}) = & \beta \text{AvgWorkerReview}_{jt} + \gamma \text{AvgFirmReview}_{jt} + \delta \text{NumJobCutPoint}_{jt} \\
 & + \rho(\text{AvgWorkerReview}_{jt} * \text{NumJobCutPoint}_{jt}) \\
 & + \tau(\text{AvgFirmReview}_{jt} * \text{NumJobCutPoint}_{jt}) \\
 & + \lambda \mathbf{X}_{it} + \mu \mathbf{J}_{it} + \nu_i + \epsilon_{ijt} \quad \mathbb{E}[\epsilon_{ijt} | \mathbf{Z}_{ijt}]
 \end{aligned} \tag{11}$$

where $\text{AvgWorkerReview}_{jt}$ is the average feedback the firm has left previously and $\text{AvgFirmReview}_{jt}$ is the average feedback workers have left the firm. $\text{NumJobCutPoint}_{jt}$ is an indicator that equals one if the firm has completed 15 or more jobs, which captures differences between new and experienced firms. \mathbf{X}_{it} are worker characteristics, \mathbf{J}_{it} are job characteristics and ν_i are worker fixed effects. Finally, \mathbf{Z}_{ijt} contains all the included covariates.

Column 1 of Table 4 presents the results. For new firms, there is very little effect of their reputation on the wages they must pay. However, for more experienced firms there is an effect of reputation. For an experienced firm, increasing the average review

it leaves by one point increases the wages it must pay by 3%. This is in contrast to the anecdotal evidence presented above, and may at first be counterintuitive. However, there is an extensive literature on firm wage effects and this is consistent with “better” firms paying higher wages and leaving better reviews (?). This suggests that we do not want to interpret these effects as causal. If the type of review that firms leave is a (fairly) permanent characteristic, then we cannot separate the causal effect of firm reputation from the effect of the type of firm without a source of exogenous variation. Column 2 presents the above specification but includes firm fixed effects. This helps control for unobservable firm characteristics and there is still a positive effect of the average review the firm leaves. However, since the firm is choosing what type of review to leave, without a source of exogenous variation, these should not be interpreted as the causal effect of reputation.

For an experienced firm, increasing the average review it receives by one point decreases the wages it must pay by 2%. This is consistent with an intuitive compensating differentials story where “worse” firms must pay higher wages. Since the worker is deciding on these reviews, we should be less worried about alternative explanations of firm types. Further, controlling for firm fixed effects increases the magnitude of the coefficient, which suggests that the reviews a firm receives do affect the wages it has to pay. While there does appear to be an effect of firm reputation, it does not appear to be in the hypothesized way above. Firms should not be worried about having to pay higher wages if they leave negative reviews.²⁵

7 Review Comments

All of the previous evidence suggests that firms are providing their private information to the labor market. However, the incentive remains for firms to gain by acting strategically. Another aspect of the review process is that firms can leave a comment along with their score. It is possible that firms use these comments to try and convince workers to work again while also leaving them less than perfect scores to bias down their market wage. For this to work, the market must place more relative weight on the review score (vs. the comment) than the worker does. One worker on the forum wondered about the prevalence of this practice:

Is it common for clients that close a contract, and rehire the freelancer on a new contract, to give less than perfect scores? Like, great written feedback, but less than perfect stars (Mariska P, 2015)?

²⁵Importantly, both firms and workers leave their reviews simultaneously and blindly, so there is no ability for retaliation based on the behavior of the other.

Table 4: Effect of Firm Reputation on Wages

		(1)	(2)
Avg. Worker Review	< 15 Jobs	0.00895 (0.00667)	0.0404 (0.0252)
	≥ 15 Jobs	0.0335** (0.00911)	0.0662* (0.0332)
Avg. Firm Review	< 15 Jobs	0.00469 (0.00652)	-0.0318 (0.0251)
	≥ 15 Jobs	-0.0220** (0.00721)	-0.0557+ (0.0336)
Worker Fixed Effects		Yes	Yes
Firm Fixed Effect		No	Yes
R^2		0.853	0.964
Number of Workers		6803	4134
Number of Firms		14413	4158
Number of Observations		24364	12176

Note: Regression contain controls for previous reviews, exam scores and participation, job titles and worker fixed effects.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses

If some firms are behaving in this way, this strategic behavior might be missed by my previous analysis. To look at the effect of review comments, I classify each comment according to its positivity using the VADER model of sentiment analysis (Hutto and Gilbert, 2014). This algorithm assigns a score between -1 and 1 based on analysis of the words in the review comment. Individual words are scored based on their positivity and intensity and the length of the comment is also taken into account. Figure 7.1 shows the overall density of the scores and Figure 7.2 shows the densities broken up by review type. There is clearly a correlation between the review score and the comment score, but there is some variation for all three review types. I then further classify the comments into Good, Bad, Really Bad, and No Comment by partitioning the (overall distribution of) comment scores in thirds.

7.1 Effects on Wages

To investigate this possible method of strategic behavior, I start by testing the relative importance of the review score and comment in the wage. If the review comment is reflected in the wage, then there is no incentive for firms to leave a glowing review comment, since it will affect the future wage. I re-run Equation 6 and interact the review types with the comment types. Recall from Table 3 that relative to a good

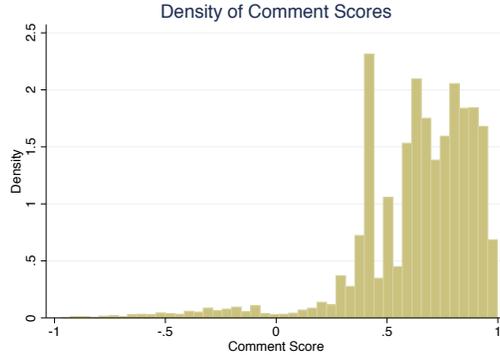


Figure 7.1: Density of Review Comment Scores

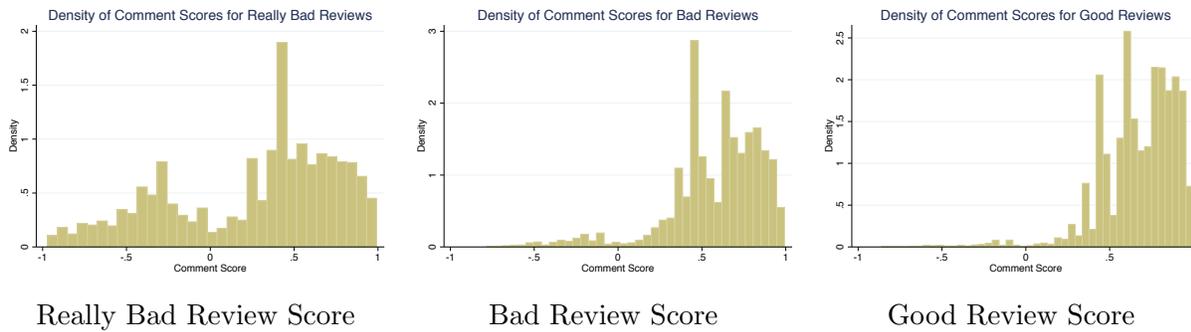


Figure 7.2: Density of Comment Scores by Review Type

review, a bad review significantly decreased the next wage. The results in Table 5 are relative to a good review with any type of comment.²⁶ Notice that receiving a bad review with a good comment also results in a significant wage decrease. Furthermore, for each review type I am unable to reject the joint hypothesis of the effects of the comments all being equal. This suggests that the review comments are largely ignored by other firms, since they do not affect wages. This means there is an incentive for a firm to leave a good comment if it will persuade the worker to work again in the future.

7.2 Effect on Contract Renewals

I now want to test whether the review comments have an effect on the worker's decision-making. To do this I re-run the contract renewal model including the interaction between review type and comment type. Figure 7.3 shows the results for workers.

There is no significant difference in workers' response to different comment types. This

²⁶I group all of the good review comments for two reasons. First, it maintains consistency between the two tables, and second, it increases power. If I compare relative to a good review with a good comment, the coefficients are similar but much less precise.

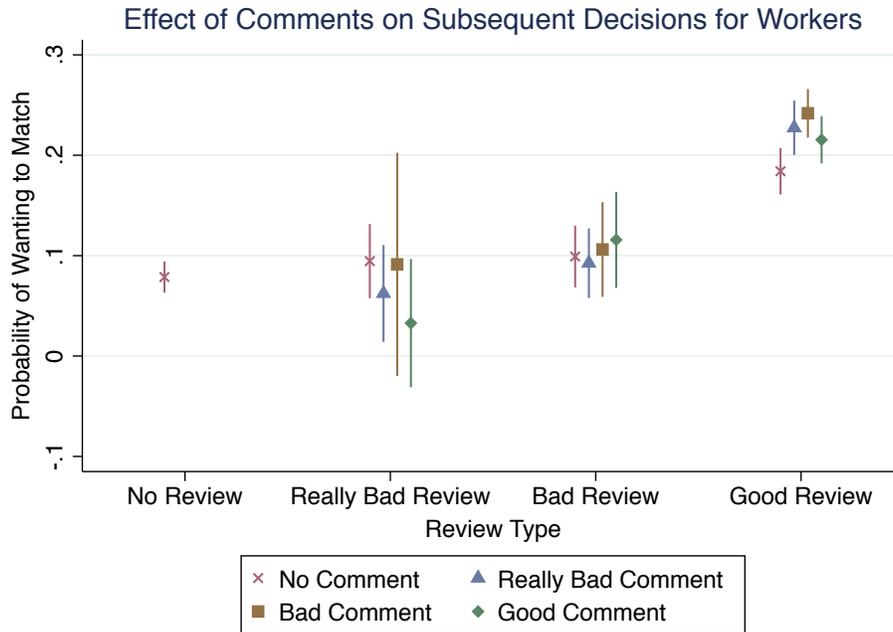
Table 5: Effect of Review Comments on Wage Changes

	(1)	(2)	(3)	(4)	
Really Bad Review	Really Bad Comment	-0.137 ⁺ (0.0787)	-0.0852 ^{**} (0.0295)	-0.0813* (0.0402)	-0.111* (0.0492)
	Bad Comment	-0.00938 (0.0902)	-0.0132 (0.0444)	-0.0296 (0.0591)	-0.0198 (0.0581)
	Good Comment	-0.202 ⁺ (0.116)	-0.0730 (0.0612)	-0.0965 (0.0670)	-0.0974 (0.0765)
	No Comment	-0.142 ^{**} (0.0415)	-0.0356 ⁺ (0.0211)	-0.0358 (0.0272)	-0.107 ^{**} (0.0322)
Bad Review	Really Bad Comment	-0.0182 (0.0399)	-0.0347* (0.0176)	-0.0316 (0.0233)	-0.0731 ^{**} (0.0277)
	Bad Comment	-0.0395 (0.0442)	-0.00555 (0.0179)	-0.00331 (0.0230)	-0.0110 (0.0326)
	Good Comment	-0.0779 (0.0494)	-0.0507 ^{**} (0.0193)	-0.0668* (0.0267)	-0.0677* (0.0321)
	No Comment	-0.0385 (0.0309)	-0.00481 (0.0133)	-0.00251 (0.0178)	-0.0351 (0.0214)
Worker Fixed Effects	Yes	No	Yes	No	
Firm Fixed Effects					
R^2	0.606	0.0449	0.129	0.431	
Number of Workers					
Number of Firms					
Number of Observations	36812	72204	69214	41577	

Note: The omitted review type is Good Review. This is showing the results for the first five jobs of a worker's career. Regressions contain controls for education, country, previous reviews, previous experiences, exam scores and participation, and job titles. The sample is the first 20 jobs of each worker's career.

⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses

suggests that workers can not be swayed by a positive review comment and that firms do not behave strategically in this manner.



Note: 95% confidence bands shown.

Figure 7.3: Effect of Comments on Subsequent Decisions for Workers

8 Conclusion

This paper investigates the information content of public, subjective, performance reviews from an online labor market. Despite incentives for firms to bias their reviews, I find that other firms are learning from these public reviews, suggesting that the reviews are positively correlated with worker-specific productivity. I investigate a number of behaviors to test this, including wage setting, exit from the market, and contract renewals, and conclude that outside firms use performance reviews to learn about the productivity of workers and that the learning happens quickly. I find that performance reviews impact whether workers stay in this online labor market: worker attrition depends on the early reviews they receive. This suggests that public performance reviews are operating on two distinct levels. They provide information to other employers about the quality of the worker, while also providing information to the worker about their ability to succeed in this labor market. I test two possible mechanisms for inducing firms to provide their private information. I find evidence that workers will not work again for a firm that leaves them a poor review or no review and that firms' overall

reputation matters. These mechanisms result in a functioning reputation system in this online labor market. These results have implications for the larger gig economy as more and more online markets rely on reputation systems to function.

References

- Agrawal, Ajay, John Horton, Nico Lacetera, and Elizabeth Lyons.** 2013. “Digitization and the Contract Labor Market: A Research Agenda.” In *Economics of Digitization*. University of Chicago Press.
- Altonji, Joseph G., and Charles R. Pierret.** 2001. “Employer learning and statistical discrimination.” *The Quarterly Journal of Economics*, 116(1): 313–350.
- Avery, Christopher, Paul Resnick, and Richard Zeckhauser.** 1999. “The Market for Evaluations.” *American Economic Review*, 89(3): 564–584.
- Benson, Alan, Aaron J. Sojourner, Akhmed Umyarov, and others.** 2015. “Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market.”
- Brown, Martin, Armin Falk, and Ernst Fehr.** 2004. “Relational Contracts and the Nature of Market Interactions.” *Econometrica*, 72(3): 747–780.
- Cabral, Luís, and Ali Hortaçsu.** 2010. “The Dynamics of Seller Reputation: Evidence from Ebay.” *The Journal of Industrial Economics*, 58(1): 54–78.
- Chen, Yubo, and Jinhong Xie.** 2008. “Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix.” *Management Science*, 54(3): 477–491.
- Chevalier, Judith A, and Dina Mayzlin.** 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research*, 43(3): 345–354.
- Cragg, John G.** 1971. “Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods.” *Econometrica*, 39(5): 829–844.
- Daniel C.** 2015. “UpWork Community Forum.”
- Dellarocas, Chrysanthos.** 2003. “The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms.” *Management Science*, 49(10): 1407–1424.
- Dellarocas, Chrysanthos, and Charles A. Wood.** 2007. “The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias.” *Management Science*, 54(3): 460–476.
- DeVaro, Jed, and Michael Waldman.** 2012. “The Signaling Role of Promotions: Further Theory and Empirical Evidence.” *Journal of Labor Economics*, 30(1): 91–147.

- Elance-oDesk.** 2014. “Elance-oDesk Global Online Work Report, 2014 Full Year Data.”
- Farber, Henry S.** 1994. “The Analysis of Interfirm Worker Mobility.” *Journal of Labor Economics*, 12(4): 554–593.
- Farber, Henry S., and Robert Gibbons.** 1996. “Learning and Wage Dynamics.” *The Quarterly Journal of Economics*, 111(4): 1007–1047.
- Fradkin, Andrey, Elena Grewal, David Holtz, and Matthew Pearson.** 2016. “Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb.” *Working Paper*.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani.** 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software*, 33(1): 1–22.
- Gneezy, Uri.** 2005. “Deception: The Role of Consequences.” *The American Economic Review*, 95(1): 384–394.
- Horton, John J.** 2010. “Online labor markets.” Springer.
- Horton, John J., and Joseph M. Golden.** 2015. “Reputation Inflation: Evidence from an Online Labor Market.” *Working Paper*.
- Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan.** 2015. “Reputation & Regulations: Evidence from eBay.” *Working Paper*.
- Hutto, C.J., and E.E. Gilbert.** 2014. “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).”
- Ipeirotis, Panagiotis G.** 2010. “Demographics of Mechanical Turk.” , (ID 1585030).
- Kahn, Lisa B.** 2013. “Asymmetric Information between Employers.” *American Economic Journal: Applied Economics*, 5(4): 165–205.
- Kahn, Lisa B., and Fabian Lange.** 2014. “Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures.” *The Review of Economic Studies*, 81(4): 1575–1613.
- Katz, Lawrence F., and Alan B. Krueger.** 2016. “The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015.” *Working Paper*.

- Lange, Fabian.** 2007. “The Speed of Employer Learning.” *Journal of Labor Economics*, 25(1): 1–35.
- Luca, Michael.** 2016. “Designing Online Marketplaces: Trust and Reputation Mechanisms.” *NBER Working Paper*.
- Luca, Michael, and Georgios Zervas.** 2015. “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud.” , (ID 2293164).
- MacLeod, W. Bentley.** 2003. “Optimal Contracting with Subjective Evaluation.” *The American Economic Review*, 93(1): 216–240.
- Mariska P.** 2015. “UpWork Community Forum.”
- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier.** 2014. “Promotional Reviews: An Empirical Investigation of Online Review Manipulation.” *American Economic Review*, 104(8): 2421–55.
- Milgrom, Paul, and Sharon Oster.** 1987. “Job Discrimination, Market Forces, and the Invisibility Hypothesis.” *The Quarterly Journal of Economics*, 102(3): 453–476.
- Mill, Roy.** 2011. “Hiring and Learning in Online Global Labor Markets.” *Working Paper*.
- Nosko, Chris, and Steven Tadelis.** 2015. “The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment.” *NBER Working Paper*.
- Pallais, Amanda.** 2013. “Inefficient hiring in entry-level labor markets.” *NBER Working Paper*.
- Petra R.** 2013. “UpWork Community Forum.”
- Pinkston, Joshua C.** 2009. “A model of asymmetric employer learning with testable implications.” *The Review of Economic Studies*, 76(1): 367–394.
- Resnick, Paul, and Richard Zeckhauser.** 2002. “Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay’s Reputation System.” In *The Economics of the Internet and E-commerce*. Elsevier.
- Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood.** 2006. “The value of reputation on eBay: A controlled experiment.” *Experimental Economics*, 9(2): 79–101.
- Scheiber, Noam.** 2015. “Growth in the ‘Gig Economy’ Fuels Work Force Anxieties.”

- Scott E.** 2016. “UpWork Community Forum.”
- Stewart, Patrick.** 2014. “Why Uber Should Let People See Their Own Passenger Ratings.” *Business Insider*.
- Tadelis, Steven.** 2016. “Reputation and Feedback Systems in Online Platform Markets.” *Annual Review of Economics*, 8(1).
- upwork.com.** 2015. “Freelancing in America: 2015 Survey.”
- Virtual working takes off in EMs.** 2012.
- Waldman, Michael.** 1984. “Job Assignments, Signalling, and Efficiency.” *The RAND Journal of Economics*, 15(2): 255.
- World Bank.** 2013. “Country and Lending Groups Data.”
- Zhu, Feng, and Xiaoquan (Michael) Zhang.** 2010. “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics.” *Journal of Marketing*, 74(2): 133–148.

A Appendix

A.1 Symmetric Employer Learning

I regress (log) wages on a number of worker characteristics, including their education, their country, and their average review score. I estimate the following equation:

$$\begin{aligned} \log(w_{it}) = & \alpha + \beta Educ_i + \gamma HighIncome_i + \delta AvgScore_{it} + \zeta g(Exp_{it}) \\ & + \eta Educ_i * g(Exp_{it}) + \theta HighIncome_i * g(Exp_{it}) \\ & + \kappa AvgScore * g(Exp_{it}) + \lambda \mathbf{X}_{it} + \epsilon_{it} \end{aligned} \quad (12)$$

where $Educ_i$ is the level of education of the worker, $HighIncome_i$ is an indicator for whether the worker is from a high income country, and $AvgScore_{it}$ is the average score of all the reviews they have received. $g(Exp_{it})$ is a function of worker experience and \mathbf{X}_{it} are other worker characteristics such as exam scores.²⁷ In the symmetric employer learning literature, the education of a worker is used as an easy to observe variable that is not strongly correlated with their productivity. However, education is self-reported on oDesk and is hard to compare across countries, so it is not obvious that the reported education of workers provides much information to employers. For this reason, I also include the country of the worker as an additional easy to observe variable. Following Altonji and Pierret (2001), I model experience as a cubic polynomial and look at the relationship between worker characteristics and wages at two points in time. Again, I limit my sample to workers who stay for at least 20 jobs.

Table 6 shows the marginal effect of worker characteristics on wages at the beginning of the worker's career and after 10 jobs. There does not appear to be any effect of having a Bachelor's degree on wages. This is potentially explained by education being self-reported by workers and thus not being a reliable signal. Initially, workers from high income countries earn 31.1% more than workers from low income countries, however, after 10 jobs the difference is cut in half. While country remains a significant factor in determining wages, its importance declines over time. Initially, the average review score is a noisy signal, but after 10 jobs, a one point increase in average review score, on a scale from 1 to 5, increases the wage by 9.3%.

²⁷For each exam on oDesk, I include an indicator for whether the worker took the exam and an indicator for whether they scored better than the median.

Table 6: Marginal Effect of Education, Country, and Reviews on Wages

	Log Wage
Bachelor's Degree	
Initial	0.0283 (0.0418)
After 10 Jobs	-0.00271 (0.0221)
High Income Country	
Initial	0.311** (0.0815)
After 10 Jobs	0.156** (0.0413)
Average Score	
Initial	0.0239 (0.0351)
After 10 Jobs	0.0934** (0.0194)
R^2	0.621
N	38291

Note: Experience is modeled as a cubic polynomial. Regression contains controls for education, previous experiences, exam scores and participation, and job titles. Sample is the first 20 jobs of a worker's career for workers who stayed at least 20 jobs.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses

A.2 A More Formal Model

Consider a model with explicit costs of leaving a review. In particular, employers choose to leave a review, $r_{it} \in [1, 5]$, by maximizing expected profit:²⁸

$$r_{it} = \arg \max_{\tilde{r}_{it}} \left(\mathbb{E}[\theta_i | s_i, r_{i1}, \dots, y_{it}] - \mathbb{E}[\theta_i | s_i, r_{i1}, \dots, \tilde{r}_{it}] \right) f(\tilde{r}_{it} | c^H) - h(\tilde{r}_{it} | c^R) - l(y_{it}, \tilde{r}_{it} | c^L) \quad (13)$$

where $f(r_{it} | c^H)$ is the probability of rehiring the worker, e.g. $f(r_{it} | c^H) = \Phi(\alpha + c^H r_{it})$ so a better review will increase the probability of rehiring the worker. $h(r_{it} | c^R)$ is an overall reputation effect, e.g. $h(r_{it} | c^R) = -c^R r_{it}$ so leaving a good review will increase expected profits. This might be because the firm is better able to attract good workers. Finally, $l(y_{it}, r_{it} | c^L)$ is a lying cost, e.g. $l(y_{it}, r_{it} | c^L) = c^L (y_{it} - r_{it})^2$ so the further the review is from the signal, the more it costs the firm. This could either be an internal cost, a firm doesn't like being dishonest, or an external cost where workers might complain about unfair reviews.

In an informative world:

$$w_{it+1} = \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it}] + \eta_{ijt+1} \quad (14)$$

$$r_{it} = \arg \max_{\tilde{r}_{it}} \left(\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, y_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}] \right) \Phi(\alpha + c^H r_{it}) + c^R r_{it} - c^L (y_{it} - r_{it})^2 \quad (15)$$

Consider four possible examples:

A.2.1 No Costs

If $c^H = c^R = c^L = 0$, then

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = - \frac{\partial \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) < 0 \Rightarrow r_{it} = 1 \quad (16)$$

So the best response of every firm is to leave the lowest possible review, which is an uninformative world. Therefore if there are no costs to leaving a review, firms have an incentive to completely bias their reviews, and so all reviews will be uninformative.

²⁸The range of reviews just needs to be some closed interval on the real line. I have chosen $[1, 5]$ for consistency with the data.

A.2.2 Infinite Rehire Costs

If $c^H = \infty$, $c^R = c^L = 0$, then

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = - \frac{\partial \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha + c^H r_{it}) + \quad (17)$$

$$\left(\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, y_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}] \right) c^H \phi(\alpha + c^H r_{it}) = 0 \quad (18)$$

$$\Rightarrow r_{it} = y_{it} \quad (19)$$

since a firm will only want to rehire a worker if $\pi \geq 0$, but a worker will only work again if they receive a perfect review. So the best response of every firm is to leave an accurate review, so all reviews are informative.

A.2.3 Infinite Reputation Costs

If $c^H = 0$, $c^R = \infty$, $c^L = 0$, then

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = - \frac{\partial \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) + c^R \Rightarrow r_{it} = 5 \quad (20)$$

So the best response of every firm is to leave a perfect review, so all reviews are uninformative.

A.2.4 Infinite Lying Costs

If $c^H = c^R = 0$, $c^L = \infty$, then

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = - \frac{\partial \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) \quad (21)$$

$$+ 2c^L \left(\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, y_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, \tilde{r}_{it}] \right) = 0 \quad (22)$$

$$\Rightarrow r_{it} = y_{it} \quad (23)$$

since there is an infinite cost to lying. So the best response of every firm is to leave an accurate review, so all reviews are informative.

A.3 Proof of Proposition 1

Proposition. *In an informative world, the marginal effect of the most recent review on the conditional mean of the change in the wage is strictly positive and decreasing with t . In an uninformative world, the marginal effect is zero.*

Proof.

In an informative world,

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}] + \eta_{ijt+1} - (\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}] + \eta_{ikt})$$

$$\frac{\partial \Delta w_{ijt+1}}{\partial r_{it}} = \frac{\partial \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}]}{\partial r_{it}}$$

we can express the conditional expectation as a linear function of the conditions:

$$\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}] = \alpha + \beta x_i + \lambda_1 r_{i1} + \dots + \lambda_t r_{it}$$

so

$$\frac{\partial \Delta w_{ijt+1}}{\partial r_{it}} = \lambda_t$$

where λ_t is the regression coefficient of r_{it} on θ_i . Since this is increasing in $Cov(\theta_i, r_{it})$ and

$$Cov(\theta_i, r_{it}) = Cov(\theta_i, g(y_{it})) = Cov(\theta_i, g(\theta_i + \epsilon_{it})) > 0$$

therefore

$$\frac{\partial \Delta w_{ijt+1}}{\partial r_{it}} > 0$$

Since λ_t is decreasing in the number of positively correlated regressors,

$$\frac{\partial^2 \Delta w_{it}}{\partial r_{it} \partial t} = \frac{\partial \lambda_t}{\partial t} < 0$$

In an uninformative world,

$$\Delta w_{ijt+1} = \mathbb{E}[\theta_i | x_i] + \eta_{ijt+1} - (\mathbb{E}[\theta_i | x_i] + \eta_{ikt})$$

$$\frac{\partial \Delta w_{ijt+1}}{\partial r_{it}} = 0$$

$$\frac{\partial^2 \Delta w_{ijt+1}}{\partial r_{it} \partial t} = 0$$

□

A.4 Proof of Proposition 2

Proposition. *In an informative world, the marginal effect of the observable characteristic on the conditional mean of the change in the wage is strictly negative and increasing with t . In an uninformative world, the marginal effect is zero.*

Proof.

In an informative world,

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}] + \eta_{ijt+1} - (\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}] + \eta_{ikt})$$

$$\frac{\partial \Delta w_{ijt+1}}{\partial x_{it}} = \frac{\partial (\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}])}{\partial x_{it}}$$

we can express the conditional expectations as linear functions of the conditions:

$$\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}, r_{it}] = \alpha^{t+1} + \beta^{t+1} x_i + \lambda_1^{t+1} r_{i1} + \dots + \lambda_t^{t+1} r_{it}$$

and

$$\mathbb{E}[\theta_i | x_i, r_{i1}, \dots, r_{it-1}] = \alpha^t + \beta^t x_i + \lambda_1^t r_{i1} + \dots + \lambda_{t-1}^t r_{it-1}$$

so

$$\frac{\partial \Delta w_{ijt+1}}{\partial x_i} = \beta^{t+1} - \beta^t$$

where β^t is the regression coefficient of x_i on θ_i . Since

$$\frac{\partial \beta^t}{\partial x_i} > 0$$

and

$$\frac{\partial^2 \beta^t}{\partial x_i \partial t} < 0$$

therefore

$$\frac{\partial \Delta w_{ijt+1}}{\partial x_i} < 0$$

and

$$\frac{\partial^2 \Delta w_{ijt+1}}{\partial x_{it} \partial t} > 0$$

In an uninformative world,

$$\Delta w_{it+1} = \mathbb{E}[\theta_i | x_i] + \eta_{ijt+1} - (\mathbb{E}[\theta_i | x_i] + \eta_{ikt})$$

$$\frac{\partial \Delta w_{it+1}}{\partial x_i} = 0$$

$$\frac{\partial^2 \Delta w_{it+1}}{\partial x_i \partial t} = 0$$

□

A.5 Wage Results Robustness Checks

To test the robustness of the choice of cut point, I rerun Equation 6 with different specifications of the variable *After*. I let the cut point be any of the first ten jobs. If the cut point is the first job, then the before group is only the first job, and the after group is all other jobs. If the cut point is the tenth job, then the before group is the first 10 jobs and the after group is everything else. Figure A.1 show the results. Really bad reviews have a strong negative effect in the first few jobs, but that effect slowly goes away. Similarly, the effect of a good review is positive over the first few reviews, and then decreases over time.

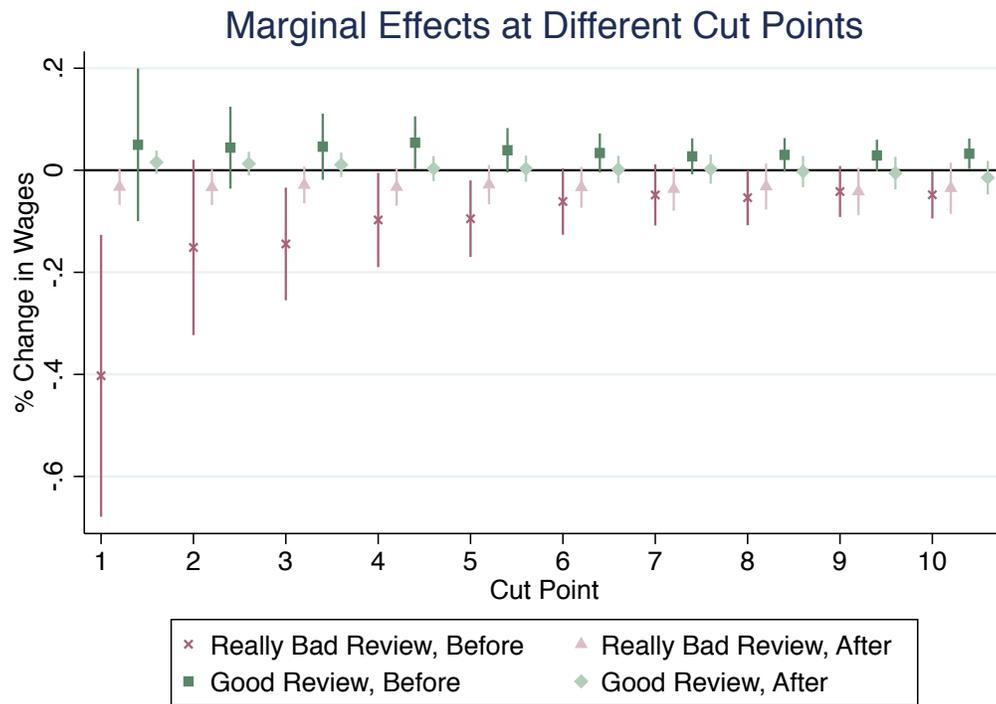


Figure A.1: Different Cut Points

I also estimate Equation 6 allowing experience to be completely non-parametric and just including dummies for each job number. Figure A.2 plots the coefficients. There is a clear trend for both a really bad and a good review, although the individual results are noisy.

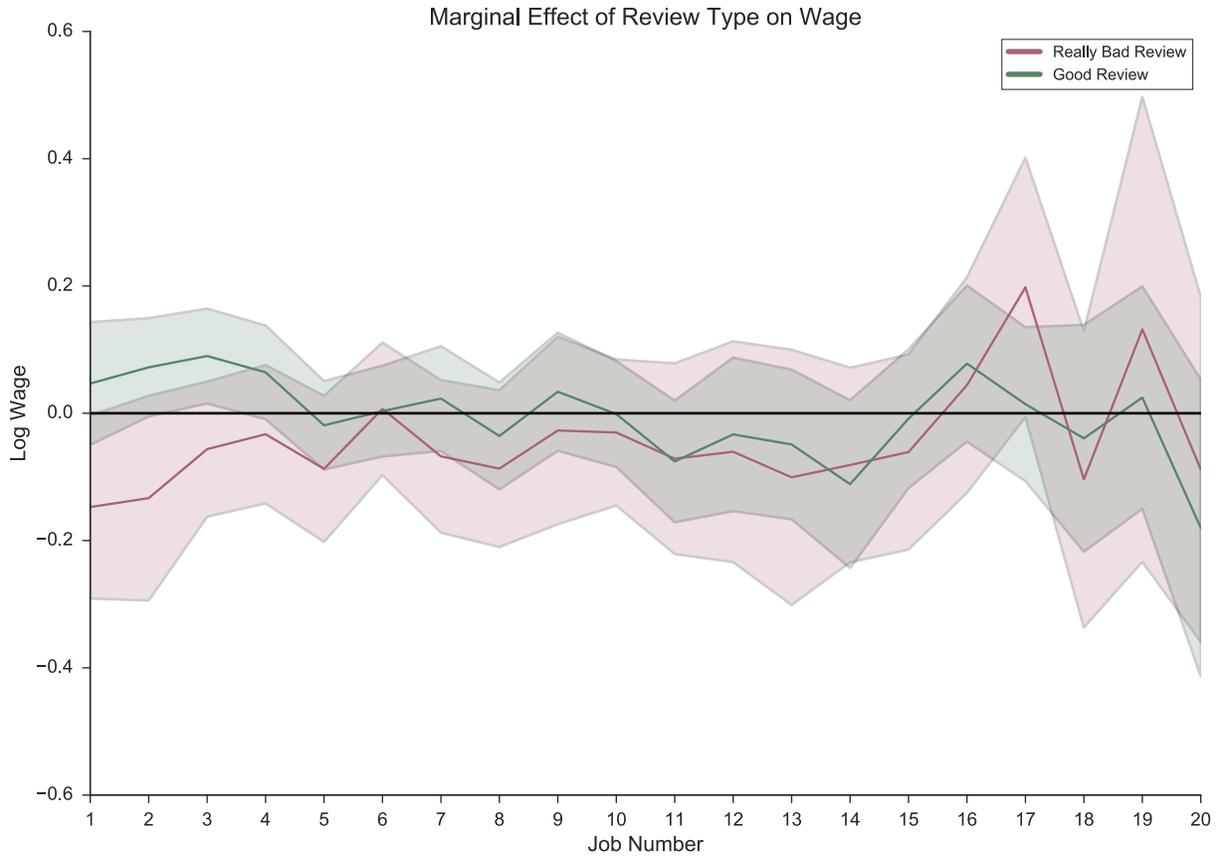


Figure A.2: Non-Parametric

A.6 Time to Job Regressions

Another way to test the informational content of the reviews is to estimate the effect of the previous review on how long it takes a worker to find another job. If workers who receive a poor review take longer to find a new job this is evidence that those reviews are providing information to the market. To do this I estimate the following model:

$$TimeToJob_{ijt} = \beta ReviewType_{it} + \lambda X_{it} + \mu J_{it} + \nu_i + \kappa_j + \epsilon_{it} \quad (24)$$

where $TimeToJob_{ijt}$ is the time, in days, between the worker's previous job and their new one. $ReviewType_{it}$ are indicators for the type of the most recent review, X_{it} are worker characteristics, such as education, exam scores, and previous review scores and J_{it} are job characteristics, which include indicators for different words that show up in the job titles. ν_i and κ_j are worker and firm fixed effects, respectively. Table 7 shows the results. There is robust evidence that a really bad review increases the time between jobs relative to a bad review and that a good review decreases the time to find a new job. This is additional evidence that these reviews are providing information to the market. Interestingly, there does not seem to be a strong effect for receiving no review. Together with the attrition and wage results, this suggests that workers who receive no review are leaving the online labor market for reasons beyond just finding another well-paying job. This is consistent with the hypothesis that workers expect to receive a review and if they do not, they become disillusioned with the platform and leave.

Table 7: Effect of Reviews on the Time to Next Job

	(1)	(2)	(3)	(4)
Really Bad Review	7.445** (1.134)	5.712** (1.058)	5.690** (1.595)	5.268** (1.519)
No Review	3.280** (0.678)	-0.579 (0.645)	2.180* (1.054)	0.0529 (1.039)
Good Review	-4.312** (0.605)	-3.371** (0.569)	-5.641** (0.935)	-4.733** (0.904)
Worker Fixed Effects	No	Yes	No	Yes
Firm Fixed Effects	No	No	Yes	Yes
R^2	0.0105	0.239	0.237	0.440
Number of Workers	21609	19849	20973	18937
Number of Firms	84751	84296	38550	38155
Number of Observations	224177	222417	177976	175553

Note: The omitted review type is Bad Review. Regressions contain controls for education, country, previous reviews, previous experiences, exam scores and participation, and job titles. The sample only includes jobs where the worker works again in the market.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses

A.7 Contract Renewal Results

Table 8: Effect of Reviews on the Probability of Matching

	Worker	Firm
No Review	0.0786** (0.00794)	0.0680** (0.00741)
Really Bad Review	0.0824** (0.0141)	0.0545** (0.0117)
Bad Review	0.101** (0.00973)	0.0801** (0.00871)
Good Review	0.217** (0.00626)	0.194** (0.00597)
Really Bad Review - No Review	0.00384 (0.0160)	-0.0135 (0.0137)
Bad Review - Really Bad Review	0.0189 (0.0170)	0.0256 (0.0144)
Good Review - Bad Review	0.116 (0.0116)	0.113 (0.0106)
N	6772	6772

Note: The bottom three results show linear combinations of the coefficients.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ Robust standard errors clustered at the worker level reported in parentheses