The Role of Reputation Systems in an Online Labor Market

Alex Wood-Doughty*

March 2018

Abstract

Much of the new "gig economy" relies on reputation systems to reduce problems of asymmetric information. There is evidence that one aspect of these reputation systems, online reviews, provide information to these markets. However, less is known about how these reviews interact and compare with other pieces of information in these markets. This paper provides a more complete picture of the reputation system in an online labor market. I compare the informational content of online reviews with other sources of information about worker ability, including the review comments, standardized exam scores, and the worker's country. I estimate the effect of each component on wages and worker attrition. Reviews have a relatively small effect on both wages and attrition, however, I am able to separate out the dual role of reviews: rewarding good workers and punishing bad ones. Finally, I investigate why firms leave reviews at all, and find that firm reputation and re-hiring considerations incentivize firms to leave informative reviews.

JEL Codes: D82 D83 J31 J49

Keywords: Online Labor Markets, Reputation Systems, Employer Learning, Performance Reviews

*Department of Economics, University of California, Santa Barbara, wooddoughty@umail.ucsb.edu. 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, the WUSTL Graduate Economics Conference, Berkeley, UCSB and Wisconsin Whitewater 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). 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. 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, you would be much less likely to get in the car of a stranger (Uber and Lyft) or send someone money for a good (Ebay). In the online labor market oDesk, the reputation system consists of public worker profiles that contain reviews left by previous employers as well as the worker's scores on various standardized exams. This paper develops a framework to understand this reputation system and allow us to quantify the importance of reviews and exams in this marketplace.

This is not the first paper to estimate the effect of oDesk reviews on wages and attrition. Pallais (2013) ran an experiment where she hired workers and left them reviews. She found that providing feedback improves their later outcomes. Stanton and Thomas (2016) estimate a similar employer learning model and find both wages and attrition are affected by reviews. The contribution of this paper is to provide a more comprehensive evaluation of the reputation system in this online labor market. This framework allows us to quantify the marginal importance of different pieces of information as well as explain how much of a worker's wage is explained by their profile. I then evaluate how important the reputation system is in differentiating workers. I also separate out two distinct roles of a reputation system–rewarding good workers and punishing bad ones–and estimate how effective this system is at achieving those goals. Finally, I investigate why firms leave reviews at all, as their feedback is critical to this system, but it is not obvious their incentives are aligned.

I start by presenting a simple theoretical framework of employer learning that relates worker characteristics to their wage. This formalizes the relationship between the various pieces of information on a worker's profile and the wage they earn. I then use this framework to estimate the marginal effect of various worker characteristics on their wage. The magnitude of these coefficients captures the importance of each characteristic in determining wages while the change in these coefficients over time can be interpreted as the speed of employer learning. I show that a worker's reviews (both the review score and comment) affect their wages and this effect increases over time. However the magnitude of this effect is small, with a one standard deviation increase in a worker's average review score only increasing their wage by 9%. I also show that a worker's country and exam scores are also important, but their effect decreases over time, consistent with the hypothesis of employer learning. To better quantify the importance of the various characteristics, I decompose the wage variance into the parts explained by each piece of information. I find that reviews explain less than 1% of the wage variation, while exams explain less than 5%. This suggests that reviews and exams convey relatively little additional information to employers that is not otherwise available.

I then turn to modeling how long workers remain in this labor market. I show that workers who do better on the standardized exams have much lower attrition in the first few jobs, but no effect later on. There is also a small consistent effect of reviews on attrition, with workers with better reviews less likely to leave. Modeling reviews non-parametrically shows that negative reviews have a bigger attrition effect than positive ones, suggesting a weeding-out effect of reviews. Combined with the results on wages, this suggests a two-part role of reviews in this market. Bad reviews are effective at removing the bottom of the worker distribution, while good reviews (and especially good review comments) reward good workers with higher wages.

Finally I turn to the question of why we see informative reviews at all. Firms who have hired a good worker have an incentive to keep this knowledge to themselves, rather then post a public review and notify their competition. This argument stems from the work of Milgrom and Oster (1987) who find that firms have an incentive to hide their talented workers by not providing public information about their skill (for example by promoting them). However, there are potentially countervailing forces that will lead firms to leave honest reviews. Benson et al. (2015) show that an employer's reputation affects their ability to hire workers, so firms in this market may be worried about reputation effects. There is also likely simple reciprocity at work, where firms want to reward good workers with a good review (Fehr and Gächter, 2000).

To investigate firm decision-making, I start by modeling a firm's decision to leave a good review as a function of different firm characteristics, controlling for the worker's characteristics (and hopefully most of their ability). I find that firms with better reviews are more likely to leave good reviews, suggesting that firm reputation may be important in this market. I also find that firms with more experience on the platform are more likely to give good reviews, which implies there is some employer learning about optimal reviewing behavior. Finally, using application data, I explicitly look at the effect of reviews on re-hiring. I find that workers who receive a good review from a firm are much more likely to want to work again for that firm, meaning that firms who have a desire to re-hire a good worker need to leave a good review. I conclude that reviews and exams are important in this online labor market, but the magnitude may be smaller than we expected. Firms are likely incentivized to leave informative reviews due to reputation and re-hiring considerations.

This paper adds to the literature on online labor markets and the role of reputation systems in these markets. As previously mentioned, Pallais (2013) and Stanton and Thomas (2016) have provided evidence that employers do learn from online reviews. Similarly, Mill (2011) and Agrawal, Lacetera and Lyons (2016) show that a worker's country is initially very important in determining wages, but that effect decreases as workers gain more experience on the platform. The current paper adds to this literature by providing a more comprehensive view of this online labor market. I confirm the previous findings on reviews and worker country while providing an estimate of the relative magnitude of these effects and comparing them to other worker characteristics. I also separate out the review score from the review comment and look at the effect of different types of reviews on both wages and attrition. This provides a more complete picture of the role reviews play in this market across the entire distribution of worker productivity. Finally, I provide some evidence for why firms leave reviews in this market. Overall this paper provides a more in-depth understanding of how online labor markets work and why they have been successful.

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 simple theoretical framework of an online labor market. Sections 4 and 5 use this model to understand the effect of worker characteristics in this market and Section 6 investigates why employers leave reviews at all.

2 Data

2.1 Online Labor Markets

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". One type of alternative work arrangement is through 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. These platforms allow people on opposite sides of the world to transact and have the potential to greatly increase welfare by reducing transaction costs.

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.[1] 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 choose to look at administrative jobs because they are more homogenous than other categories and the majority of them are hourly jobs.[2] I observe every job these 18,147 workers have done

[1] Workers 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.

[2] There 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.

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. Table 1 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.[3] 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.[4] The median (partial) career is 15 jobs with the median job being a fairly significant time commitment: over a week long and over 40 hours. 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 job at a time, so their career on oDesk is fairly sequential. To focus on the early careers of workers, I will limit my sample to the first 20 jobs of each worker's career.

The two main sources of information that are unique to this type of market are standardized exams and public performance reviews. For the standardized exams, I focus on the five exams that were taken by at least 5% of the workers. Table 2 provides summary statistics for the exams. The first column gives the percentage of workers who took that exam, while the second gives the median score (scaled between 0 and 1).

For the public performance reviews, I will focus on the average review score of the worker. This score is the average of the worker's previous review scores (each between 1 and 5). This information is prominently displayed on the worker's profile and is likely seen by every firm that is interested in hiring the worker. Figure 2.1 shows the density of average review scores. It is highly skewed towards a perfect score.

I will also utilize the text from a worker's previous reviews. When a firm leaves a review of a worker, they also have the option of leaving written feedback on the worker. To look at the effect of these 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. I then standardize the comment scores to be on a scale between 1 and 5 to match the review scores and

[3] I classify countries according to the World Bank Country Income Classification (World Bank, 2013).
[4] However, the hourly wage is comparable to the average hourly wage in India and the Philippines.

Table 1: Summary Statistics for Workers

High Income Country: OECD	0.16 (0.37)
High Income Country: Non-OECD	0.012 (0.11)
Upper Middle Income Country	0.064 (0.24)
Lower Middle Income Country	0.61 (0.49)
Low Income Country	0.15 (0.36)
Bachelor's Degree	0.49 (0.50)
Master's Degree	0.080 (0.27)
Number of Outside Experiences	2.72 (2.61)
Number of oDesk Exams	0.88 (1.05)
Average Wage ($)	5.91 (4.55) [4.34]
Average Job Duration (Hours)	97.3 (180.4) [36.2]
Average Job Duration (Days)	18.0 (29.2) [9.46]
Average Career Length (Jobs)	23.9 (32.8) [14]
Average Number Jobs at One Time Conditional on Having a Job	1.46 (2.14)
Number of Workers	19598

Note: Standard deviations in parentheses, medians in brackets.

Table 2: Summary Statistics for Exams

	Took Exam	Median Score
English Spelling	0.54	0.75
		(0.23)
English Basic Skills	0.47	0.65
		(0.23)
English Vocabulary	0.24	0.66
		(0.23)
Office Skills	0.22	0.65
		(0.24)
Email Etiquette	0.17	0.59
		(0.25)
N	19598	

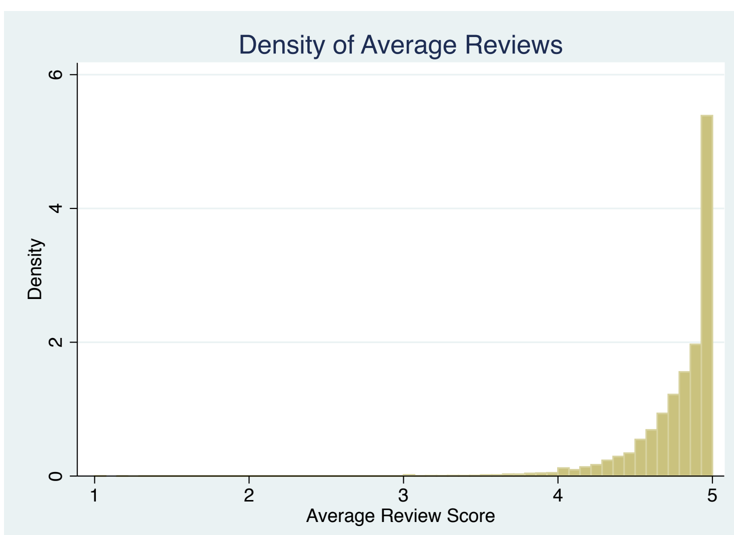


Figure 2.1: Distribution of Average Reviews

take the average of a worker's previous comments. Figure 2.2 shows the density of the average comment scores. While the comments are still overwhelmingly positive, there is more of a distribution.[5] While the review comments are less prominently displayed on a worker's profile, it still seems likely that an interested firm would at least briefly peruse a worker's review comments. For the majority of my analysis I will sum the worker's average review score and their average comment score to generate an average combined score.

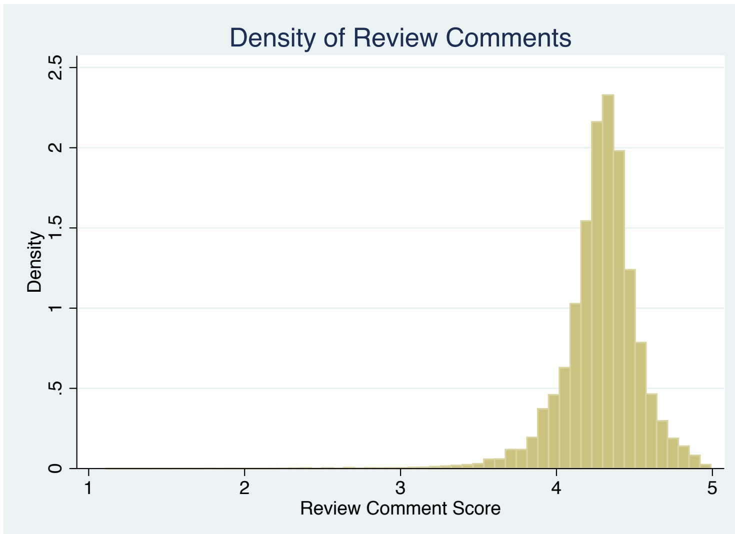


Figure 2.2: Distribution of Average Comments

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 (θ_i) is a linear function of their observable characteristic and an unobservable, permanent, idiosyncratic component (ν_i): $\theta_i = \gamma x_i + \nu_i$.[6] 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}. For simplicity, I will assume that firms honestly reveal their signal: $r_{it} = y_{it}$. More generally, the results hold for review setting subject to $r_{it} = g(y_{it})$ for some weakly monotonically increasing function g with at least one strict increase. In this competitive marketplace, wages are a function of the expected productivity of the worker, conditional on the information

[5]Note that the VADER model is not imposing any kind of distribution on the data, the algorithm simply scores each comment independently.

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

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.

If we assume joint normality of the unobservable components[7]:

$$\begin{pmatrix} \nu \\ \epsilon \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & 0 \\ 0 & \sigma_\epsilon^2 \end{pmatrix} \right) \tag{1}$$

then we can write the wage for worker i in period $t+1$ as:

$$w_{it+1} = \lambda_t \bar{r}_{it} + (1-\lambda_t)\gamma x_i + \eta_{ijt+1} \tag{2}$$

where

$$\bar{r}_{it} = \frac{1}{t} \sum_{\tau=1}^{t} r_{i\tau} \tag{3}$$

and

$$\lambda_t = \frac{t\sigma_\nu^2}{t\sigma_\nu^2 + \sigma_\epsilon^2} \tag{4}$$

λ_t measures the contribution of a worker's average review to their wage and therefore $\frac{\partial}{\partial t}\lambda_t$ can be interpreted as the speed of learning (through reviews) about worker productivity.[8] This framework lends itself to a standard linear model where we compare the effect on wages of different worker characteristics over time.

4 Wage Results

I regress (log) wages on a number of worker characteristics, including their country, eduction, exam scores, and their average review score. I also include controls for firm and job characteristics. I estimate the following equation:

$$\begin{aligned} \log(w_{ijt}) = \alpha + \boldsymbol{\beta X_{it}} + \gamma g(Exp_{it}) + \boldsymbol{\delta X_{it} * g(Exp_{it})} \\ + \boldsymbol{\lambda Z_{it}} + \boldsymbol{\kappa W_{jt}} + \boldsymbol{\mu J_{ijt}} + \epsilon_{ijt} \end{aligned} \tag{5}$$

where $\boldsymbol{X_{it}} = [HighIncome_i, Bachelor_i, \boldsymbol{Exam_{it}}, AvgScore_{it}]$. $HighIncome_i$ is an indicator for whether the worker is from a high income country, $Bachelor_i$ is an indicator for whether the worker lists a bachelor's degree, $\boldsymbol{Exam_{it}}$ are the worker's exam scores for each standardized exam, and $AvgScore_{it}$ is the average score over all the reviews they have received. Following Altonji and Pierret (2001), $g(Exp_{it})$ is a cubic polyno-

[7]Again, this is not necessary for the basic results to hold, however it makes the math much cleaner.
[8]This idea is formalized in Lange TODO

mial of worker experience, $\boldsymbol{Z_{it}}$ are other worker characteristics such as previous wages and resume experience, $\boldsymbol{W_{jt}}$ are firm characteristics such as experience and average wage paid, and $\boldsymbol{J_{it}}$ are job characteristics, which include indicators for different words that show up in the job titles.[9]

Figure 4.1 plots the marginal effects of the worker characteristics over time. For both the exams scores and the review scores, I have standardized the distributions to be mean zero and standard deviation one, so we can interpret the effects as a one standard deviation increase. The displayed effect for exams is the sum of the five effects, so the interpretation is the effect of a worker increasing all five of their exam scores by a standard deviation. For the combined review score, I plot the effect of increasing both a worker's review score and comment score by a standard deviation.

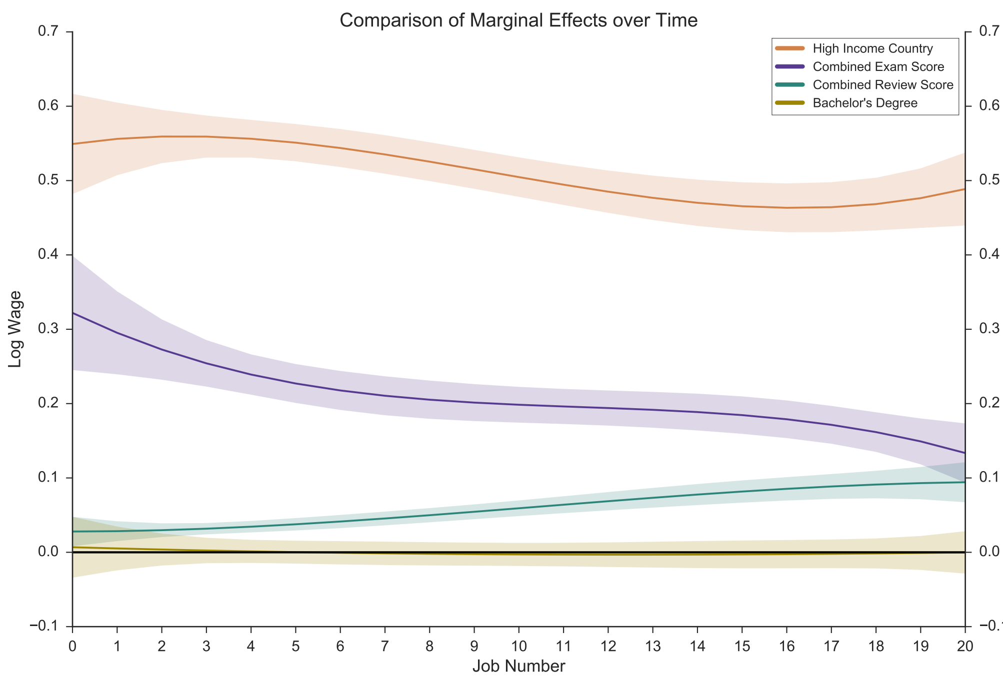


Figure 4.1: Marginal Effects of Worker Characteristics on Wages

The importance of the worker's country and their exam scores in determining wages decreases over time, while the effect of the average review score increases. This suggests

[9]To 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.

there is learning about worker productivity through reviews. There is no effect of a Bachelor's degree in this market. This is potentially explained both by education being self-reported by workers and that most workers are not from the United States, so interpreting the importance of education is more difficult. Initially, workers from high income countries make on average 55% more than workers from middle income countries, however this wage gap decreases to 48% after 20 jobs. Similarly, for the first job, increasing all five exam scores by a standard deviation, increases a worker's wage by 32%, but this drops to only 13% after 20 jobs. The wage effect of reviews increases over time, from only a 3% increase for a one standard deviation increase in the first job, to a 9% bonus in the 20th.

These results highlight a few important facts. First, there is evidence of employer learning from reviews in this market. The importance of good reviews increases over the course of a worker's career. Further, there is learning from both the review score and the comment score. Figure 4.2 shows that the two scores follow very similar trajectories over a worker's career. While there is evidence of significant learning from reviews, the overall effect on wages is much less than that of the worker's country and exam scores. Section 4.1 has a more detailed comparison of the relative importance of different characteristics.

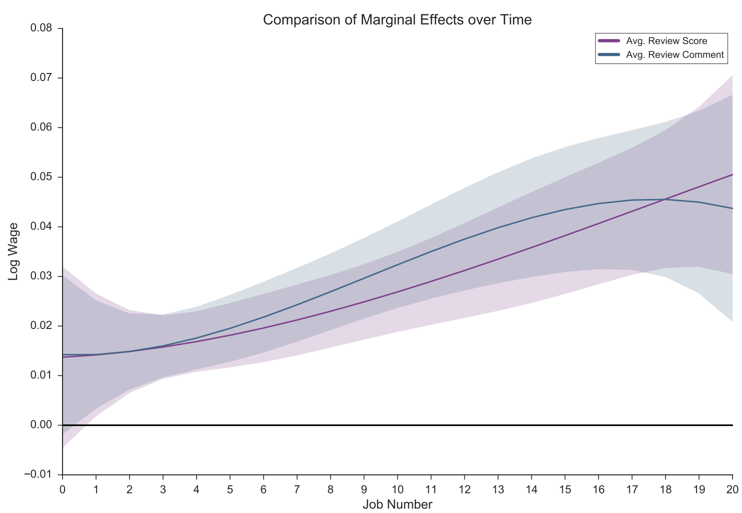


Figure 4.2: Marginal Effects of Reviews on Wages

While Figure 4.1 shows learning from performance reviews, it is assuming a linear relationship in the combined review score. However, the density of average reviews is distinctly non-Gaussian, so there may be important non-linearities in the effects of reviews on wages. To attempt to capture these non-linearities, I first separate out the combined score into the average score and the comment score and then group each score into three quantiles. Figure 4.3 shows the marginal effects over time of the good and bad quantiles relative to the middle quantile. Unsurprisingly, there is a monotonic relationship between the three quantiles, with better reviews leading to higher wages. For the review score, there does not seem to be much important non-linearity. However, for the comment scores, bad comments have a strong negative effect early in a worker's career, while good comments are more important later on. This relationship make intuitive sense. Early in a worker's career, they only have a few reviews, so a bad comment will likely be read by potential employers and can have a detrimental effect. As we will see in Section 5, reviews also play a weeding-out role, so conditional on making it to 20 jobs, bad review comments do not have as much of a negative effect. However an employer might be willing to pay a premium for a worker with really good comments.

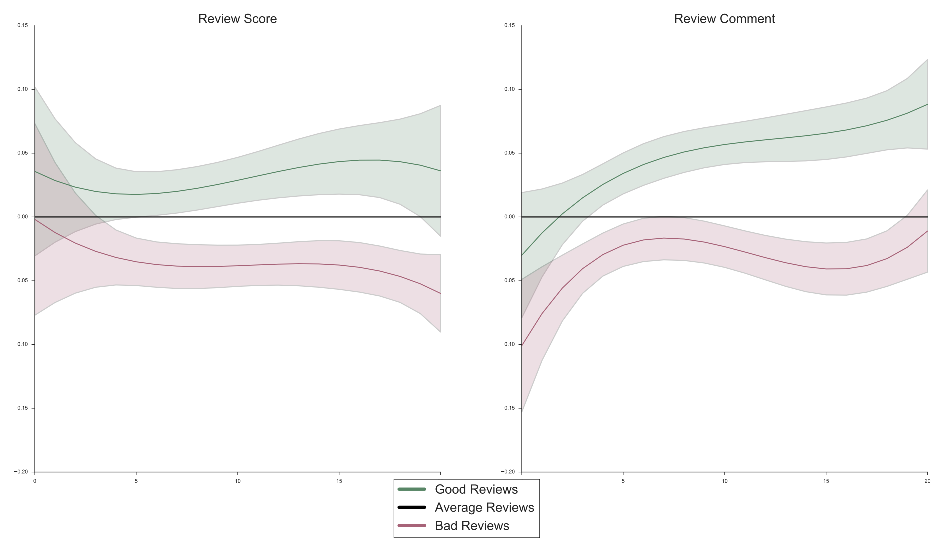


Figure 4.3: Marginal Effects of Review Quantiles

4.1 Decomposition of Wage Variance

Another way to compare the relative importance of different worker characteristics is to decompose the variance in wages into different components. This allows for an easy comparison between worker characteristics about which contribute the most to the observed distribution of wages. Table 3 shows the results for three different specifications. The first columns is the standard specification given in Equation 5. The second and third specifications add firm and worker fixed effects respectively. This allows us to see how the worker characteristics change as we control for more of the variation across workers and firms. The worker's country of origin and type of job are the biggest explainers of wages. The combined effect of the exams accounts for 5% of wage variation while the combined review score accounts for less than 1%. While there is evidence of employer learning from these reviews, this suggests that they are not providing that much information. Adding in firm fixed effects reduces the impact of the various worker characteristics, suggesting that there is some sorting among workers and firms. If some firms are more picky (and therefore only hire workers with good reviews) but pay more we would see a positive correlation between reviews and wages without firm fixed effects, but this effect would decrease once we controlled for firm heterogeneity. Finally, the third column adds worker fixed effects. This absorbs more of the effects of reviews and exams, since they are relatively time-invariant. This specification lets us see that worker characteristics are by far the most important in determining wages, followed by firm characteristics and then job characteristics. In total these components explain 91% of the wage variation. These results suggest that reviews do matter in determining wages, but they explain only a very small part of wage variation. Another way in which reviews might matter on this platform is by helping to determine which workers stay in the labor market.

Table 3: Decomposition of Wage Variance

	Contributions (%)					
	(1)		(2)		(3)	
Total	100		100		100	
Predicted Effects of $X'\beta$	60.7		37.8		12.3	
Education		0.023		-0.010		
Worker Country		11.6		10.3		
Avg. Review Score		0.29		0.25		0.18
Avg. Review Comment		0.34		0.27		0.17
Exams		4.71		3.63		1.10
Number of Jobs		-0.32		0.45		1.84
Outside Work Experience		0.24		0.27		
Job Characteristics		20.9		22.6		9.01
Firm Characteristics		22.9				
Firm Fixed Effects			40.3		23.8	
Worker Fixed Effects					54.9	
Residual	39.3		21.9		9.00	
Number of Workers	19598		17982		14247	
Number of Firms	57540		22808		21817	
Number of Observations	122045		87313		82644	

Note:

5 Attrition Results

In the previous section, we saw that reviews have a small effect on a workers' wages. It is possible that these public performance reviews also have an effect on worker attrition: whether they continue working in this online labor market. Since reviews have an effect on workers' wages, a bad review may cause the worker's expected next wage to fall below their reservation wage, and they may instead choose to leave the online labor market. Similarly, the review (and comment) may signal to the worker that they are not a good fit for this type of work and cause them to seek out other forms of employment. To test the hypothesis that public performance reviews affect worker attrition in an online labor market, I model workers' careers using a proportional hazard model. My outcome of interest is whether that worker is hired again on oDesk. This outcome is a function of a cubic baseline hazard rate as well as worker characteristics. In particular, I allow the hazard rate to depend on all the variables from Equation 5 and letting the effects vary with time. Thus, the hazard rate for worker i, firm j, and job t is given by:

$$\lambda_{ijt} = \lambda_0(t) * exp\left(\boldsymbol{\beta_t X_{it}} + \boldsymbol{\delta Z_{it}} + \boldsymbol{\kappa W_{jt}} + \boldsymbol{\mu J_{ijt}}\right) \tag{6}$$

where $\boldsymbol{X_{it}} = [HighIncome_i, Bachelor_i, \boldsymbol{Exam_{it}}, AvgScore_{it}]$ and $\boldsymbol{\beta_t}$ vary with t.

To investigate the effect of different worker characteristics on attrition, Figure 5.1 plots the marginal effects on the hazard rate for each characteristic of interest. There is basically no effect of a worker's country of their attrition rate. Workers who do well on the exams are much less likely to leave after the first few jobs, however conditional on staying for a few jobs, exam scores no longer have any effect on attrition. Finally, the combined review score has a constant effect on attrition over the worker's career. A one standard deviation increase in a worker's combined review score reduces their probability of exit by 1%.

Table 5.2 splits up the combined review score and shows that the score itself is generally more important for attrition, but that comments do have a significant effect.

Finally, I re-run the hazard model using the quantiles of the review scores. Table 5.3 shows the results. For both the review score and the comment score, in the first few jobs negative reviews have a bigger effect than positive reviews. This suggests that reviews on oDesk have a weeding-out effect, with workers who receive bad reviews leaving the market. This also means the wage effects of reviews in the previous section are likely attenuated as we only observe the wages of workers who stay in the market. Together, these results 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 try to understand firms' decisions to leave informative reviews.

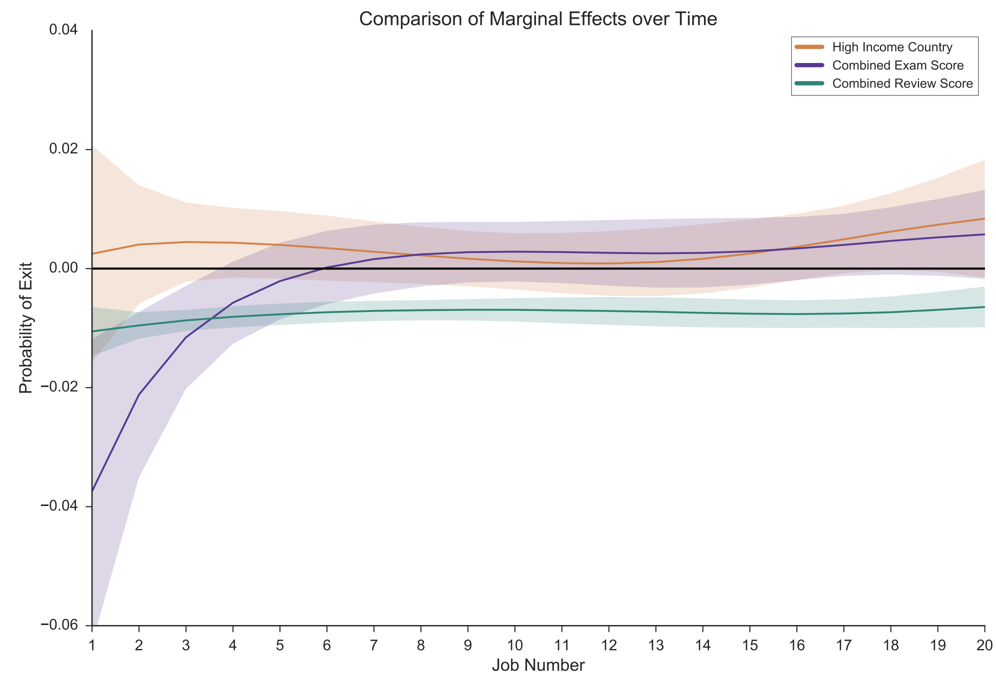


Figure 5.1: Effect of Worker Characteristics on Attrition

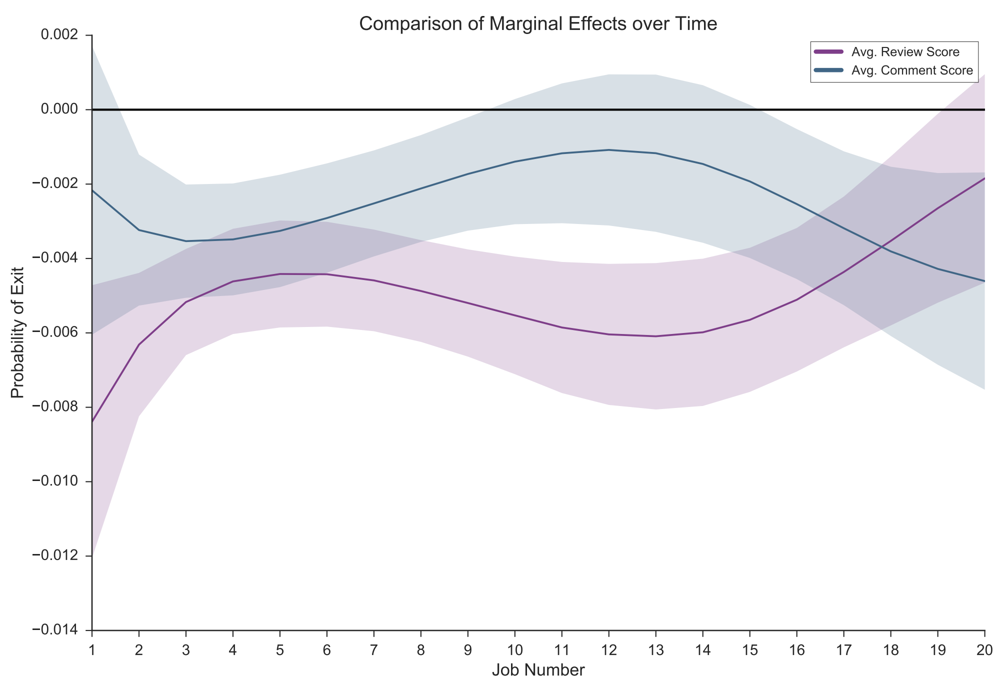


Figure 5.2: Effect of Review Scores on Attrition

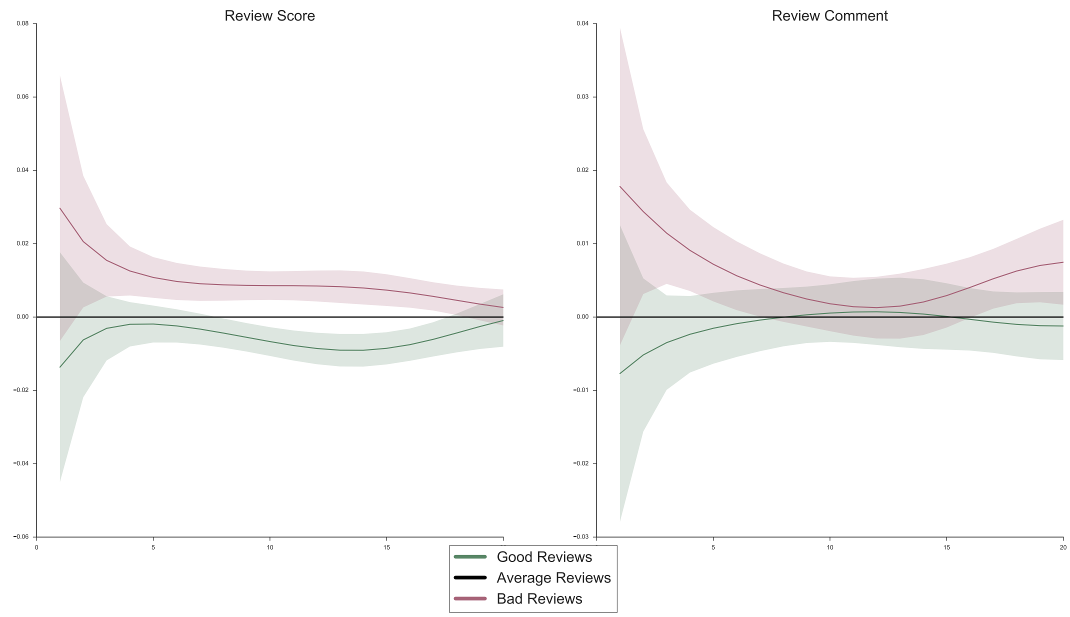


Figure 5.3: Effect of Review Quantiles on Attrition

6 Why Do Employers Leave Reviews?

As pointed out in the Introduction, there are likely costs and benefits to a firm choosing to leave a good review of a worker. As such, there is no clear a priori prediction on firm behavior. However, the wage and attrition results suggest these reviews are providing information to the marketplace. In this section I model firm behavior with respect to review-leaving to try to understand why firms are leaving informative reviews. I start by modeling the decision to leave a good review as a function of worker and firm characteristics. This purely correlative approach helps us understand which types of firms leave good reviews.

I model the binary decision to leave a good review (a perfect review score) as a function of the variables in Equation 5, while also controlling for the wage the worker received for the job. In particular, the firm characteristics I focus on are experience (number of jobs in this marketplace), average firm review score, average wage paid (all split into quintiles) and the country of the firm. I also look at the duration of the job and the number of applicants to see how those affect the probability of a good review. While this regressions is almost certainly capturing some selection by firms to hire different types of workers, by conditioning on the observable characteristics of the worker (their country, exams, reviews, etc.), this approach does provide some insight into the relationship between firm characteristics and review-leaving behavior.

Figure 6.1 plots the relationship between the various firm and job characteristics and the probability of a good review. Firm experience and review score both have a fairly linear relationship with review-leaving. Firms with more experience and better reviews are more likely to give good reviews. Low-paying firms are more likely to leave good reviews than higher paying firms. This is potentially because they have lower expectations. Similarly, firms from lower income countries are more likely to leave good reviews than those from higher income countries. Job duration also has a relationship with the probability of good reviews. Short jobs are more likely to have good reviews than longer ones. This again could be because of expectations. Finally, jobs with the lowest number of applicants are the most likely to have good reviews. This is consistent with firm re-hiring, where firms post jobs to hire a specific worker. I will investigate this behavior in more detail in the next section.

The main takeaways from these results are that expectations matter, not all jobs are similar and the probability of good reviews reflects that fact. More interestingly, there is a relationship between both firm experience and firm reputation (as measured by the average review score) and the probability of leaving a good review. This is potentially explained by firms learning over time about the importance of reviews to the marketplace.[10]

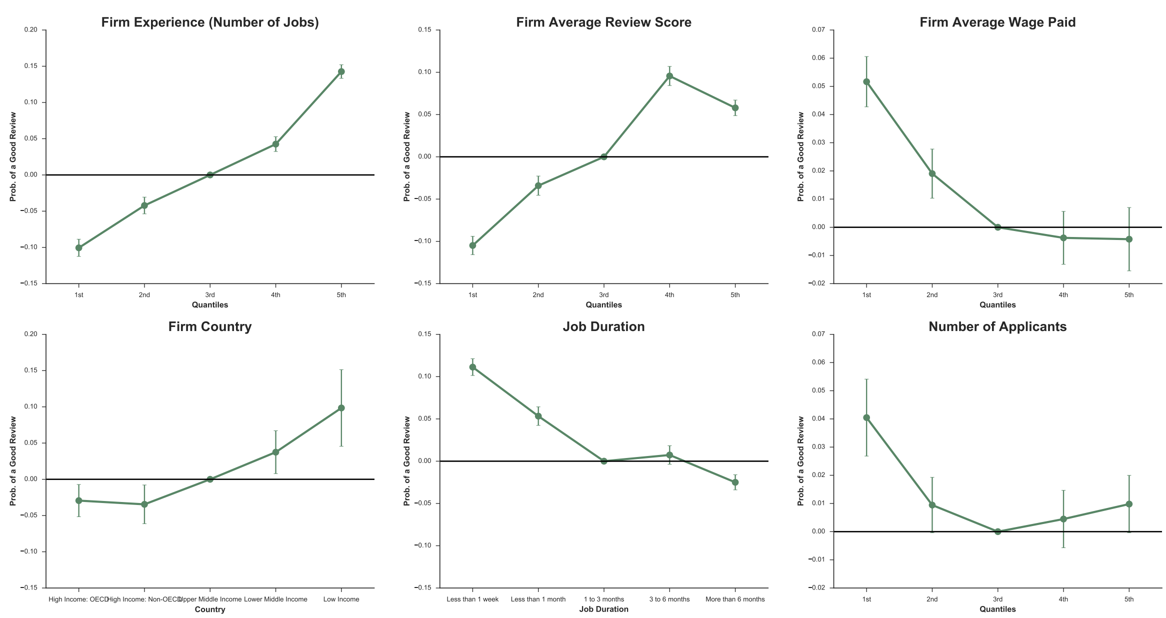


Figure 6.1: Effect of Firm Characteristics on Good Reviews

I also look at the main worker variables and their effect on the probability of a good

[10] Worker reviews and firm reviews are left simultaneously and blindly, so the review effect is not just a simple case of tit-for-tat.

review. Once again, these are interacted with a cubic in experience, to track the effects over time. These results are conditioning on the wage the worker earned for that job, meaning we can interpret these effects of worker characteristics as how likely they were to do a "good" job given that they were hired. Figure 6.2 plots the results. There is no significant effect of worker country on good reviews. So while the worker's country is strongly correlated with the wages that they earn, conditional on those wages, there is no additional effect of country. There are significant effects of exam scores and review scores on the probability of a good review. This suggests that these scores do provide valuable information about the worker. Not only do these scores help workers get hired for better jobs, for the same job, workers with good scores are more likely to do a good job than those without.

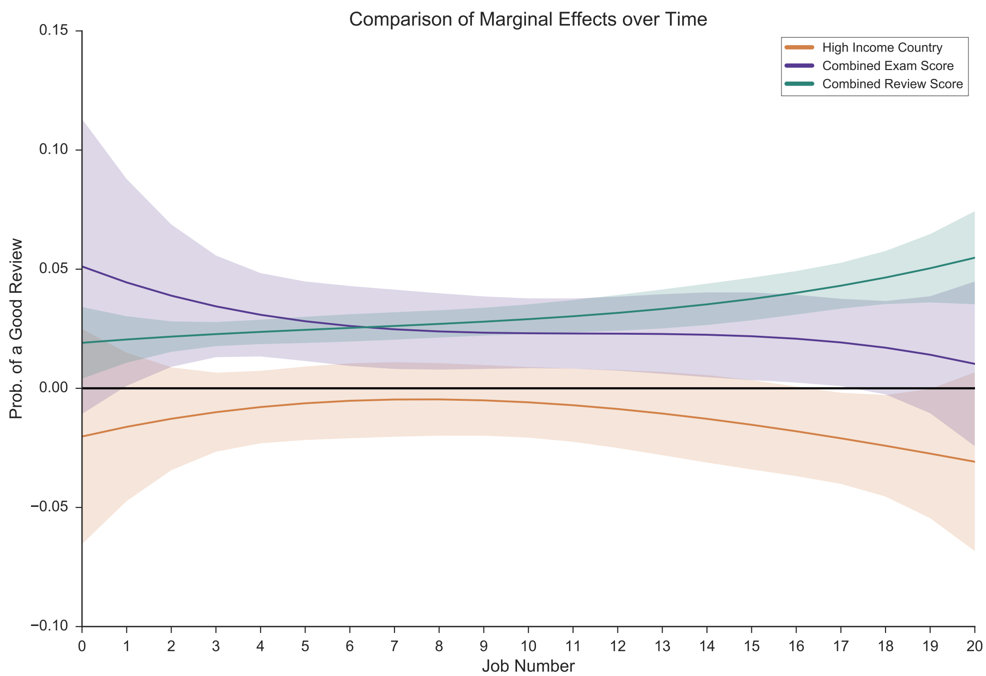


Figure 6.2: Effect of Worker Characteristics on Good Reviews

6.1 Effect of Reviews on Subsequent Decisions

The previous section highlighted some relationships between firm characteristics and the probability of leaving a good review. Another reason to leave a good review is if you want to re-hire that worker. In online labor markets there are a relatively high percentage of repeat transactions (around 20% in my sample). Given the uncertainty about the quality of a worker, firms may be inclined to re-hire workers they know are good. However, in order to be able to re-hire a worker, a firm may need to provide them with positive feedback.

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 positive feedback. This suggests a mechanism by which firms are forced to leave positive feedback.[11] To test this mechanism, I turn to application data that provides information on worker decision-making. 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 the characteristics of their first encounter.

6.2 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.[12] 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

[11] Firms 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.

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

job.[13] 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.[14]

6.3 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 6.3 shows the decision tree.

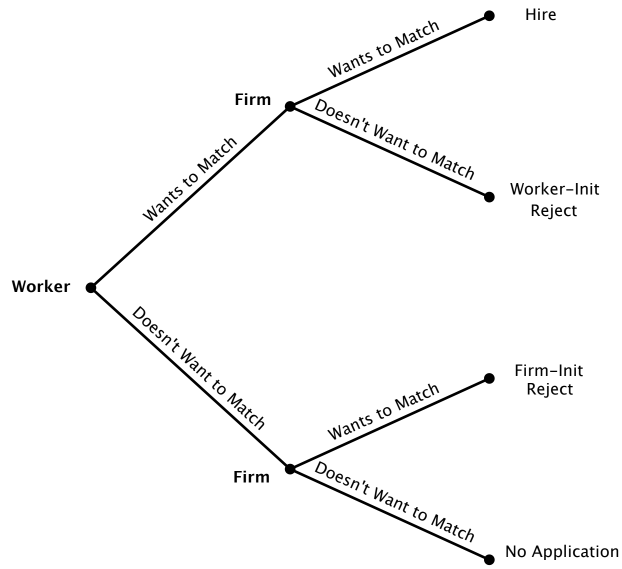


Figure 6.3: Decision Tree for Contract Renewals

If both worker and firm decide to match, I observe an application and a hiring. 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 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:

$$WantToMatch = \mathbb{1}\{x'\beta + \epsilon_w > 0\} \tag{7}$$

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

[14]If 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.

while the firm's decision is:

$$WantToMatch = \mathbb{1}\{z'\gamma + \epsilon_f > 0\} \tag{8}$$

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 5 for both the workers' and firms' decisions. The likelihood function for the full model is given in Equation 9.

$$f(outcome|\beta, \gamma) = \begin{cases} \Phi(x'\beta)\Phi(z'\gamma) & \text{if Hired} \\ \Phi(x'\beta)(1 - \Phi(z'\gamma)) & \text{if Worker App, Reject} \\ (1 - \Phi(x'\beta))\Phi(z'\gamma) & \text{if Client App, Reject} \\ (1 - \Phi(x'\beta))(1 - \Phi(z'\gamma)) & \text{if No App} \end{cases} \tag{9}$$

where Φ is the standard normal CDF.

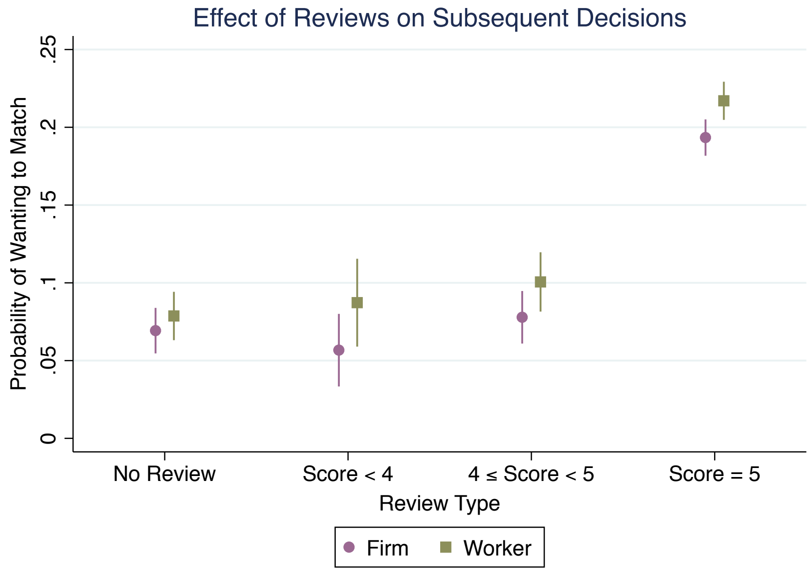
6.4 Results

Figure 6.4 shows predicted contract renewal rates based on maximum likelihood estimates of Equation 9.[15] Both firms and workers are twice as likely to want to match again if the previous review was perfect than if it was any other score. 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. These estimates provide evidence for a mechanism that induces firms to leave informative reviews. Together with the correlative results, this suggests firms leave good reviews for re-hiring and reputation purposes and this behavior is learned over time.

7 Conclusion

Online review scores have received a fair amount of attention in the literature, especially with the rapid expansion of the gig economy. To date, these studies have focused on the relationship between reviews and worker outcomes, without considering the role of other information. This paper provides a more comprehensive investigation of the entire

[15] Table 4 in Appendix A.4 provides the coefficients.



Note: 95% confidence bands shown.

Figure 6.4: Effect of Reviews on Subsequent Decisions

reputation system in an online labor market. Reviews are important, but their relative magnitude is small compared to standardized exams and the worker's country. This means that reviews do provide signal about a worker's ability, and there is employer learning from them, but they do not provide as much information as other signals. As these markets continue to grow, understanding the relationship between different signals and how the platform can improve these signals becomes more important. These results suggest that the reputation system in this market is functional, but there is significant room for improvement.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

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.

Agrawal, Ajay, Nicola Lacetera, and Elizabeth Lyons. 2016. "Does standardized information in online markets disproportionately benefit job applicants from less developed countries?" *Journal of International Economics*, 103: 1–12.

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."

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2012. "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Science*, 59(2): 265–285.

Brown, Jennifer, and John Morgan. 2006. "Reputation in Online Auctions: The Market for Trust." *California Management Review*, 49(1): 61–81.

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." https://community.upwork.com/t5/Coffee-Break/Scratching-my-head/m-p/144963.

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.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159–181.

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*.

Hu, Nan, Jie Zhang, and Paul A. Pavlou. 2009. "Overcoming the J-shaped Distribution of Product Reviews." *Commun. ACM*, 52(10): 144–147.

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." https://community.upwork.com/t5/Coffee-Break/Scratching-my-head/m-p/144932.

Masterov, Dimitriy V., Uwe F. Mayer, and Steven Tadelis. 2015. "Canary in the e-Commerce Coal Mine: Detecting and Predicting Poor Experiences Using Buyer-to-Seller Messages." *EC '15*, 81–93. New York, NY, USA:ACM.

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." https://community.upwork.com/t5/Clients/Contractors-who-DEMAND-5-stars/m-p/27531.

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." https://community.upwork.com/t5/Clients/Ratings-go-from-1-to-5/m-p/177396.

Stanton, Christopher, and Catherine Thomas. 2017. "Information Frictions and Observable Experience:The New Employer Price Premium in an Online Market." *Working Paper.*

Stanton, Christopher T., and Catherine Thomas. 2016. "Landing the First Job: The Value of Intermediaries in Online Hiring." *The Review of Economic Studies*, 83(2): 810–854.

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."

Xu, Haitao, Daiping Liu, Haining Wang, and Angelos Stavrou. 2015. "E-commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service." *WWW '15*, 1296–1306. Republic and Canton of Geneva, Switzerland:International World Wide Web Conferences Steering Committee.

Zervas, Georgios, Davide Proserpio, and John Byers. 2015. "A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average." Social Science Research Network SSRN Scholarly Paper ID 2554500, Rochester, NY.

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 Additional Wage Results

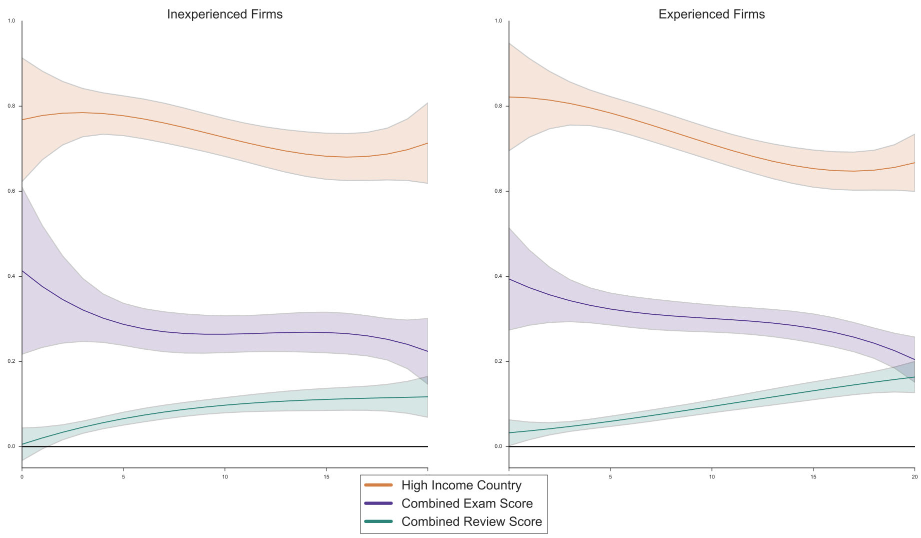


Figure A.1: Split by Firm Experience

A.2 Additional Attrition Results

A.3 Additional Review-Leaving Results

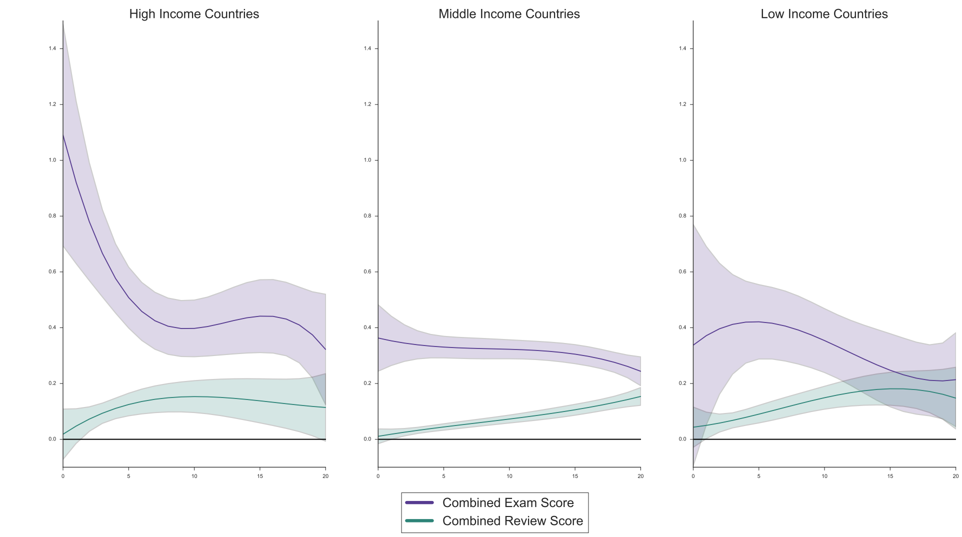


Figure A.2: Split by Worker Country

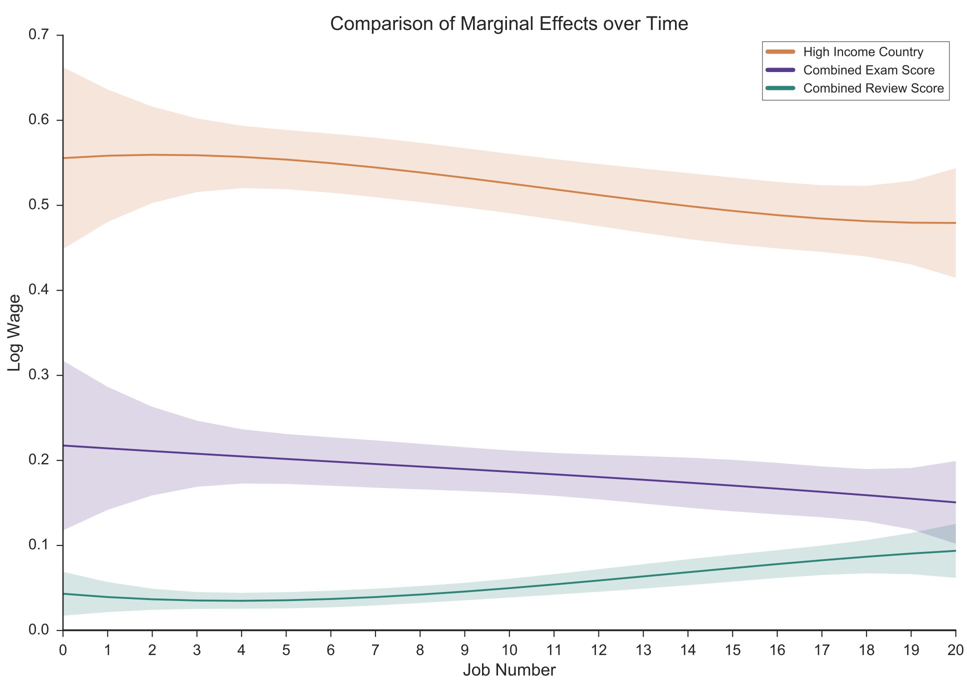


Figure A.3: With Firm Fixed Effects

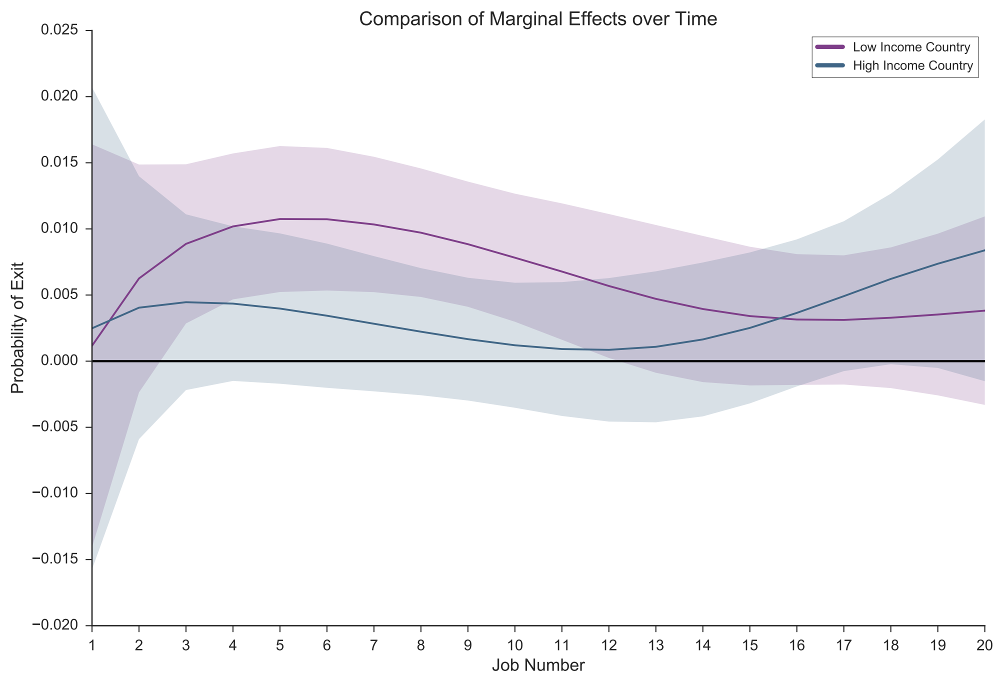


Figure A.4: Effect of Country on Attrition

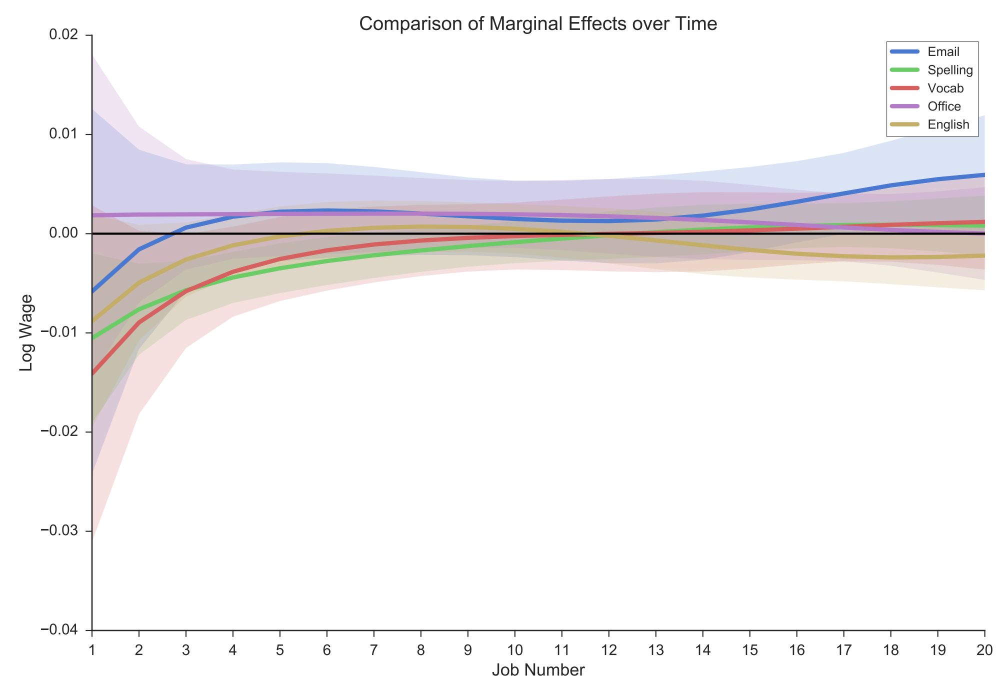


Figure A.5: Effect of Exams on Attrition

A.4 Contract Renewal Results

Table 4: Effect of Reviews on the Probability of Matching

	Worker	Firm
No Review	0.0787**	0.0693**
	(0.00793)	(0.00747)
Really Bad Review	0.0872**	0.0567**
	(0.0144)	(0.0119)
Bad Review	0.101**	0.0778**
	(0.00973)	(0.00861)
Good Review	0.217**	0.193**
	(0.00626)	(0.00596)
Really Bad Review - No Review	0.00855	-0.0125
	(0.0163)	(0.0139)
Bad Review - Really Bad Review	0.0133	0.0211
	(0.0172)	(0.0145)
Good Review - Bad Review	0.116 **	0.116**
	(0.0116)	(0.0105)
N	6779	6779

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