Title
Three Essays in Economics

Author
Wood-Doughty, Alex

Publication Date
2017

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Three Essays in Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Alex Wood-Doughty

Committee in charge:

Professor Peter Kuhn, Co-Chair
Professor Douglas Steigerwald, Co-Chair
Professor Richard Startz

June 2017

The Dissertation of Alex Wood-Doughty is approved.

Professor Richard Startz

Professor Douglas Steigerwald, Co-Chair

Professor Peter Kuhn, Co-Chair

June 2017

Three Essays in Economics

Acknowledgements

First and foremost, I thank Brenna for all her support throughout the trials and tribulations of the past five years. I also thank my family for their continuing encouragement and support and for instilling in me a passion for education.

I will be forever indebted to my committee for all their help along the way. Peter Kuhn read countless drafts of my papers and provided amazing feedback every time. Doug Steigerwald made sure I saw the big picture and offered invaluable advice. Dick Startz was always there to chat about research or life and always provided encouragement. I thank them not only for their research advice and support, but also for their friendship and collegiality throughout my time at UCSB.

I also thank Ted Bergstrom, Clément de Chaisemartin, and members of the Econometrics Research Group and Labor Lunch for their feedback and advice.

Curriculum Vitæ
Alex Wood-Doughty

Education

2017	Ph.D. Economics, University of California, Santa Barbara
2013	M.A. Economics, University of California, Santa Barbara
2012	B.A. Economics and Political Economy, Washington University in St. Louis

Research Fields

Labor Economics and Econometrics

Committee Members

Peter Kuhn (Co-Chair), Douglas G. Steigerwald (Co-Chair), Richard Startz

Grants and Fellowships

2012-2017	University of California Regents Fellowship
2016	NET Institute Summer Grant

Presentations

2016-2017	NET Institute Conference, University of Wisconsin, Whitewater, WUSTL Economics Graduate Student Conference, California Econometrics Conference, Berkeley Labor Lunch
2015-2016	WEAI Annual Conference, Trans-Pacific Labor Seminar, ZEW Workshop on Atypical Employment

Teaching Assistant

2016-2017	Ph.D. Econometrics
2014-2016	Introduction to Econometrics
2013-2014	Introduction to Economics

Abstract

Three Essays in Economics

by

Alex Wood-Doughty

In this dissertation I study a new emerging labor market as well as a very old question about education. In all three chapters, I collect and build complex datasets and combine formal theoretical modeling with advanced econometric techniques. This allows me to answer interesting questions about peer effects in education, as well as employer learning and competitiveness in online labor markets.

In Chapter 1, Richard Startz and I propose an improved estimator of peer effects using network data. The ability to estimate peer effects in network models has been advanced considerably by the IV model of Bramoullé, Djebbari and Fortin (2009). While such IV estimates work well for very sparse networks, they exhibit very weak power for networks of even modest densities. We review and extend the findings of Bramoullé, Djebbari and Fortin (2009) and then propose an alternative estimator. We show that our new estimator works approximately as well as IV in very sparse networks and performs much better in networks of moderate density. To highlight the benefits of our proposed estimator, we provide an empirical application where we estimate peer effects in individual schools.

In Chapter 2, I study whether employers learn from public, subjective, performance reviews. 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, subjec-

tive, 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.

In Chapter 3, I analyze the competitiveness of online labor markets. A number of new labor market innovations have promised a more competitive marketplace. This paper tests this claim by comparing an online labor market with traditional labor markets. First, using data from an online labor market, I estimate a two-way fixed effects model and compare my estimates to the existing literature on traditional labor markets. Then, using a model of imperfect competition, I relate the estimated employer-specific wage premiums with the structural parameters governing market competitiveness. Combined with reasonable assumptions about differences in these labor markets, I show that the estimated differences in the variance of employer-specific wage premiums imply differences in competitiveness. My results suggest that online labor markets are more competitive than traditional labor markets.

Contents

List of Figures

List of Tables

Chapter 1

Improved Estimation of Peer Effects using Network Data

1.1 Introduction

Estimation of peer effects in networks has undergone enormous growth. Much of this growth is due to the breakthrough work of Bramoullé, Djebbari and Fortin (2009) (hereafter BDF) who show that if we know the network structure–who interacts with whom–and there is enough sparsity of the network, peer effects are identified. Much of the peer-effect literature is framed in the Manski (1993) linear-in-means framework, i.e. my outcome depends linearly on the average outcomes of my friends. Manski shows that, without knowing anything about the group structure, the linear-in-means model is not identified. BDF make two contributions. First, they show that the additional information available from knowing the structure of network interactions often permits identification. Second, they offer an instrumental variable strategy for estimating such models.

While BDF prove that identification holds for almost all networks, their simulations show that meaningful inference is only possible in very sparse networks. For networks with density above 5%, (the average person is connected with 5% of the network), their instrumental variables estimation offers very limited power. Our primary contribution is to offer a nonlinear estimator that we show has enormously improved power in denser networks with only a small departure from nominal size. We also extend and confirm BDF's Monte Carlo results on bias and precision, and further frame the issues in the weak instrument framework. (See Nelson and Startz (1990*a*,*b*) and Staiger and Stock (1997))

Much of the practical concern about network density arises in the study of friendship networks in school, which are typically relatively dense within a classroom or a single school. BDF use their technique to identify peer effects of recreational activities by middle and high-schoolers. Their technique has also been used in a number of other papers: De Giorgi, Pellizzari and Redaelli (2010) look at peer effects in college major choice, Lin (2010) looks at academic achievement, and Fortin and Yazbeck (2015) look at obesity. Calvó-Armengol, Patacchini and Zenou (2009) use a similar method to investigate peer effects in school performance. All of these studies use data on the friendship networks in schools to identify the peer effects.

While some schools are sufficiently large to ensure a low density friendship network, there are many with densities above 5%. To get around this problem, the previous studies generate a sparse network by stacking multiple school networks and assuming there is no interaction across schools. To illustrate the increased power of our proposed method, we give an empirical example in which we identify peer effects without stacking across schools.

The rest of the paper is organized as follows. Section 2 describes the linear-in-means model and explains the common method of estimating peer effects. Section 3 provides some Monte Carlo simulations that highlight the problems with the common method for higher density networks. Section 4 proposes an alternative estimation strategy and Section 5 provides simulations that highlight its benefits. Section 6 applies the new method to the Add Health data and estimates school-level peer effects.

1.2 Linear-in-Means Model

Consider the following model, taken from BDF:

$$y_i = \alpha + \beta \frac{1}{n_i} \sum_{j \in g_i} y_j + \gamma x_i + \delta \frac{1}{n_i} \sum_{j \in g_i} x_j + \epsilon_i \tag{1.1}$$

$$\epsilon_i | X \sim iidN(0, \sigma_\epsilon^2)$$

where g_i is the set of friends of person i and n_i is the size of g_i. We model person i's outcome, y_i, as a simple average of her friends' outcomes, y_j, an observable characteristic, x_i, and an average of her friends' observable characteristics, x_j. Following Manski (1993), we will call β the endogenous social effect and δ the exogenous social effect. The endogenous social effect, or peer effect, can be interpreted as the relationship between an average friend's outcome and your outcome. In a classroom setting, we might be interested in the peer effect of test scores: how your average friend's test score affects your own score. We assume $|\beta| < 1$, as is standard. The exogenous social effect measures the influence of your friends' characteristics on your outcome. If we believe that the education level of a student's parents affects their achievement in the classroom, we might also believe the education level of their friends' parents has an effect. Importantly, we want to separately identify the endogenous social effect from the exogenous effect. They are both capturing the influence of peers, however they have very different policy implications. In particular, in many situations the endogenous social effect creates a multiplier on changes in the exogenous variables.

Following BDF, we can stack the observations and write the model as:

$$Y = \alpha\iota + \beta GY + \gamma X + \delta GX + \mathcal{E} \tag{1.2}$$

$$\mathcal{E}|X \sim N(0, \sigma^2_{\mathcal{E}} I_n)$$

where G is the adjacency matrix, a row-normalized $n \times n$ matrix that describes the network and ι is an $n \times 1$ vector of ones. The adjacency matrix starts as a matrix of ones and zeros where $G_{ij} = 1$ if i is friends with j. Each row is then normalized by its sum so both the endogenous and exogenous social effects can be interpreted as the effect of an average friend. Since this is a simultaneous system, in order to separately identify both β and δ we need an instrument. To see this, we can rewrite Equation 1.2 to give the reduced form by collecting terms on Y:

$$\begin{aligned} Y &= \alpha\iota + \beta GY + \gamma X + \delta GX + \mathcal{E} \\ &= (I_n - \beta G)^{-1}(\alpha\iota + \gamma X + \delta GX + \mathcal{E}) \end{aligned} \tag{1.3}$$

We can write $(I_n - \beta G)^{-1}$ as a series expansion, $\sum\limits_{k=0}^{\infty} \beta^k G^k$, so:

$$
\begin{aligned}
Y &= \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \gamma X + \delta GX + \mathcal{E}) && (1.4)\\
&= \sum_{k=0}^{\infty} \beta^k G^k (\gamma X + \delta GX) + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) && (1.5)\\
&= \gamma X + \sum_{k=1}^{\infty} \beta^k G^k \gamma X + \sum_{k=0}^{\infty} \beta^k G^k \delta GX + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) && (1.6)\\
&= \gamma X + \sum_{k=1}^{\infty} \beta^k G^k \gamma X + \sum_{k=1}^{\infty} \beta^{k-1} G^k \delta X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) && (1.7)\\
&= \gamma X + \sum_{k=1}^{\infty} (\beta^k \gamma + \beta^{k-1}\delta) G^k X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) && (1.8)\\
&= \gamma X + \sum_{k=1}^{\infty} \beta^{k-1} (\beta\gamma + \delta) G^k X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) && (1.9)\\
&= \gamma X + (\beta\gamma + \delta) GX + \beta(\beta\gamma + \delta) GGX + \sum_{k=3}^{\infty} \beta^{k-1} (\beta\gamma + \delta) G^k X + \frac{\alpha}{1-\beta}\iota + \sum_{k=0}^{\infty} \beta^k G^k \mathcal{E} && (1.10)
\end{aligned}
$$

If we only use information on X and GX, we cannot separately identify $[\alpha, \beta, \gamma, \delta]$. However, since GGX does not show up directly in Equation 1.2 but it does show up in the reduced form of GY, BDF propose that the characteristics of the friends of friends (GGX) can be used to instrument for the friends' outcomes, GY, and Equation 1.2 can be estimated via 2SLS. The structural parameters are identified as long as GGX is not perfectly collinear with ι, X, and GX.[1] Another way of saying this is that there must be some intransitive triads: second-order friends who are not also first-order friends. The instrument must be providing some information about GY. While this is true for most networks, the amount of information decreases as the network gets more dense.

[1]Generally, for any $p > 1$, $G^p X$ is a valid instrument for GY as long as $G^p X$ is not perfectly collinear with $[\iota, X, GX]$.

As everyone has more friends, everyone's second-order friends start to look more and more similar. We show below that "weak identification" in this context results in biased standard errors but only modestly biased coefficients. In other words, the classical IV asymptotic distributions are about right although the estimated distributions are not. The more important issue, correctly identified by the asymptotic distributions, is that the estimator is of very low power.

1.2.1 Estimation

Once we have an instrument, estimation proceeds via 2SLS.[2] Following BDF, we assume there are network-specific unobservables, α. To control for these unobservable characteristics, we apply a within-transformation by multiplying each term by $I_n - \frac{1}{n}(\iota\iota')$. Note, that this transformation changes the identification requirement slightly. As proven in BDF, now the structural parameters are identified as long as $GGGX$ is not perfectly collinear with $[X, GX, GGX]$.

Define the following:

$$R = I_n - \frac{1}{n}(\iota\iota') \tag{1.11}$$

$$W = [RGY, RX, RGX] \tag{1.12}$$

$$Z = [RX, RGX, RGGX] \tag{1.13}$$

$$P = Z(Z'Z)^{-1}Z' \tag{1.14}$$

$$\theta = [\beta, \gamma, \delta]' \tag{1.15}$$

[2]Lee, Liu and Lin (2010) adopted spatial models to allow estimation via maximum likelihood. They show their method is more efficient than 2SLS, however they are still relying on variation in GGX for identification.

Then the 2SLS estimator is:

$$\hat{\theta}_{2SLS} = (W'PW)^{-1}W'PRY \tag{1.16}$$

and the asymptotic variance is:

$$Var(\hat{\theta}_{2SLS}) = \sigma^2_{\mathcal{E}}(W'PW)^{-1} \tag{1.17}$$

where σ^2 is estimated by

$$\hat{\sigma_{\mathcal{E}}}^2 = \frac{1}{n}((RY - W\hat{\theta})'(RY - W\hat{\theta})) \tag{1.18}$$

1.3 2SLS Simulation Results

To examine the properties of this estimator, we run Monte Carlo simulations. We generate data from known parameters and then estimate the model in an attempt to recover those parameters. We will vary the density of the network to see how that affects the estimation. We follow BDF by setting the network size, $N = 240$, $\alpha = 0.7683$, $\beta = 0.4666$, $\gamma = 0.0834$ and $\delta = 0.1507$. We draw X from a log-normal distribution with mean 1 and variance 3. For each simulation, we generate a random reciprocal network with the probability of any link equal to the density. We then row normalize so the weight on each individual's friends sums to 1. We also draw the error vector, $\mathcal{E} \sim N(0, 0.1)$. From there, we can generate the outcome Y according to Equation 1.3. We simulate 10,000 networks with each of the following densities: 0.01, 0.025, 0.05, 0.1, 0.2, 0.3.

Figure 1.1 plots the distribution of the estimated $\hat{\beta}$s. As in the BDF findings, the estimates become imprecise around a density of 5% and there is no meaningful inference for dense networks, in that the distributions become essentially flat. Note also some increase in bias as the density increases.

We are mainly interested in inference about $\hat{\beta}$, so we construct the following test statistic: $t_{\hat{\beta}} = \frac{\hat{\beta}-0.4666}{SE_{\hat{\beta}}}$ and compare the statistic to a 5 percent critical value. Figure 1.2 plots the empirical size of this test. At very low densities, the test achieves approximately the correct size. However, in more dense networks the test is notably undersized.

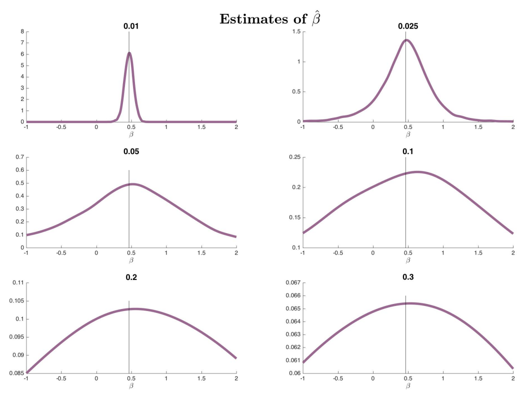


Figure 1.1: Distribution of $\hat{\beta}$ for varying network densities

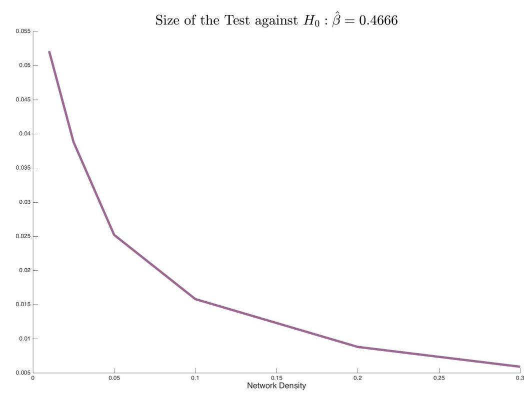


Figure 1.2: Size of $\hat{\beta}$ for varying network densities

To test the power, we let the true β vary uniformly between $[0, 1]$ and then test the hypothesis that $\hat{\beta} = 0.4666$. Figure 1.3 plots the power of the test at the specified densities.

Even for the 5% network there is limited power. For densities of 10% or larger, power effectively equals size, meaning there is no information in the test statistic. To investigate the cause of this low power, we look at the first stage regression:

$$RGY = \pi_1 RX + \pi_2 RGX + \pi_3 RGGX + V \tag{1.19}$$

Figure 1.4 plots the median F-statistic for the coefficient on the instrument, π_3. This shows that the instrument is very weak for the 5% density network and closely mirrors Figure 1.3.

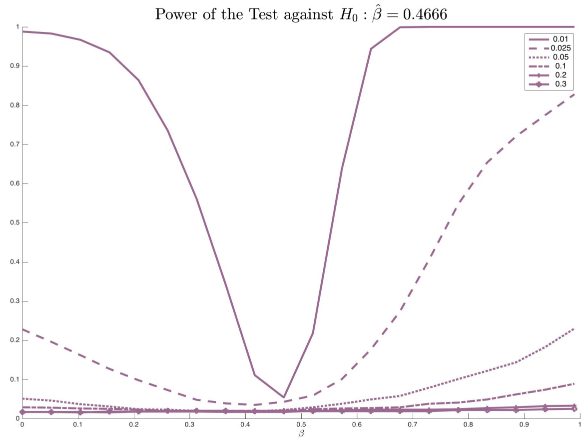


Figure 1.3: Power

The normal cause of a weak instrument problem is that the chosen instrument is only slightly correlated with the endogenous variable (Nelson and Startz, 1990*a*). Here, the reason the instrument is so weak is because GGX is highly collinear with $[\iota, X, GX]$. Figure 1.5 plots the log of the median condition number for $W'PW$. The condition number is a measure of collinearity, and numbers above 3.4 (log(30)) are evidence of high multicollinearity (Belsley, Kuh and Welsch, 1980). Here we can clearly see the relationship between collinearity, the weak instrument, and the lack of power of the estimates. The reason we are getting weak identification for higher density networks is because GGX is not providing much new information.

1.3.1 Anderson-Rubin Test

To further investigate the weak instrument hypothesis, we can run an Anderson-Rubin (1949) test, which should have exact finite sample size up to Monte Carlo error. Figure 1.6a plots the size of the Anderson-Rubin test. As expected the test has the proper 5% size across all densities. Figure 1.6b plots the power of the Anderson-Rubin test by varying the true β uniformly between $[0, 1]$ and setting $\beta_0 = 0.4666$. This shows the well

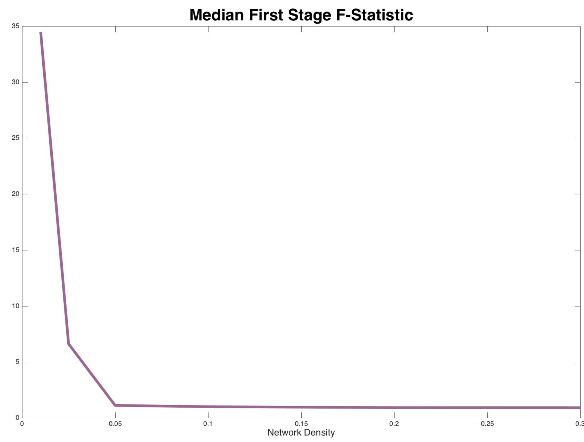


Figure 1.4: First Stage F-Statistic

known fact that the Anderson-Rubin test has limited power. However, a comparison to Figure 1.3 shows that the power of the Anderson-Rubin test is not much worse than the power of 2SLS. This, together with the multicollinearity results, suggests that the difficulties with instrumental variable inference in denser networks reflect low information content rather than a failure of the usual asymptotic approximations.

1.3.2 Size with True Error Variance

To investigate the causes of the test being undersized, we recompute the standard errors using the true error variance, σ^2, in place of the estimated error variance. Figure 1.7 plots the size of the test for both standard 2SLS and using the true σ^2. When we use the true error variance, the test has proper size, suggesting that the reason 2SLS is undersized is that it is overestimating the error variance. Thus the asymptotic distribution gives a good approximation to the empirical distribution with $N = 240$.

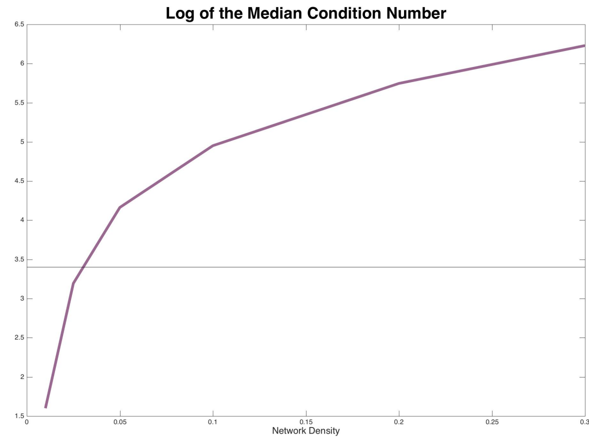
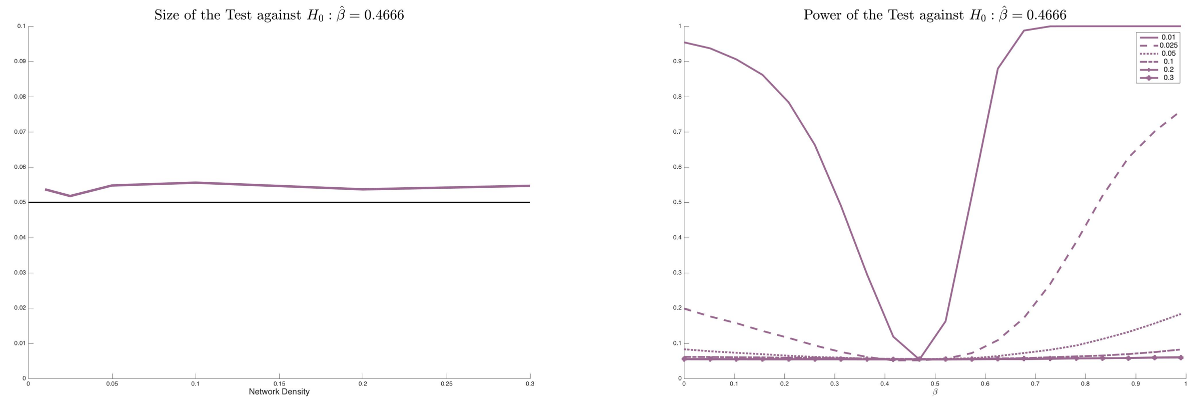


Figure 1.5: Log Condition Number

1.3.3 Other Sample Sizes and Number of Variables

While choosing $n = 240$ and a single covariate is consistent with BDF, we might also be interested in the behavior of the estimator in different situations. In particular, we might like to know how it performs in larger samples and with a more realistic set of covariates. To illustrate this, we run simulations with $n = 1000$ and with simulated covariates that are similar to the empirical application of BDF.[3] Figure 1.8 shows the results for the three sparse denisities (1%, 2.5%, 5%). There is a marked decline in power when including a number of other covariates, as well as increasing the sample size. This suggests that the performance of the 2SLS estimator depends on more than just the density of the network, and that for empirical applications, even a 1% density network may not be enough to guarantee reasonable power.

[3]We simulate the covariates to approximate the summary statistics in Table 1.1.



(a) Size

(b) Power

Figure 1.6: Anderson-Rubin Test

1.4 Alternative Estimation Strategy (Residual Series Expansion Estimation)

As seen in the previous section, any estimation strategy that relies on variation in GGX will have difficulty with dense networks. To avoid this, we exploit another part of Equation 1.10 to identify β:

$$\begin{aligned} Y &= \gamma X + (\beta\gamma + \delta)GX + \beta(\beta\gamma + \delta)GGX \\ &\quad + \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)G^k X + \frac{\alpha}{1-\beta}\iota + \sum_{k=0}^{\infty} \beta^k G^k \mathcal{E} \\ &= \gamma X + (\beta\gamma + \delta)GX + \beta(\beta\gamma + \delta)GGX \\ &\quad + \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)G^k X + \frac{\alpha}{1-\beta}\iota \\ &\quad + \mathcal{E} + \beta G\mathcal{E} + \beta^2 G^2 \mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E} \end{aligned} \tag{1.20}$$

Notice that β shows up in the series expansion of the error term. The goal of our strategy is to estimate β from this series expansion. To do this, we must first get a good

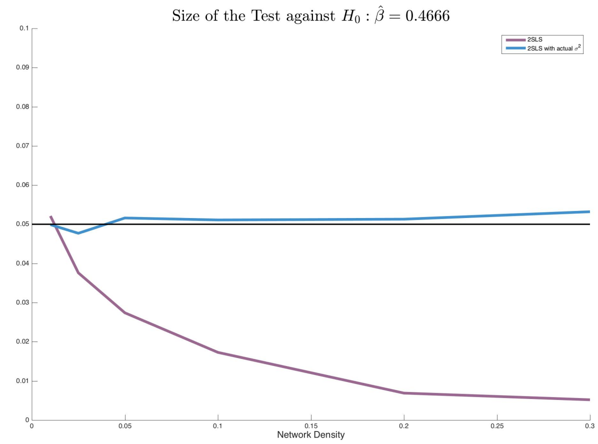


Figure 1.7: Size with True Error Variance

approximation of the error term. Since $\mathcal{E}$ is uncorrelated with X, the regression of Y on $[\iota, X, GX, GGX]$ will generate the best linear predictor of Y.[4] As we have already seen, the omitted higher order terms are highly correlated with the included terms. Because the higher order terms are projected onto the included terms, and also because powers of β fade toward zero, the residuals will approximate the series expansion with $\mathcal{E}$.

$$Y = \lambda_0 \iota + \lambda_1 X + \lambda_2 GX + \lambda_3 GGX + \eta \tag{1.21}$$

$$e = Y - \hat{Y} = \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)(G^k X - \mu_{0k} - \sum_{l=1}^{3} \mu_{lk} G^{l-1} X) \tag{1.22}$$

$$+ \mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E}$$

$$\approx \mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E} \tag{1.23}$$

where the μ coefficients come from the projection of the higher order terms onto the

[4]Any subset of $G^p X$ will also work, see Subsection 1.4.1

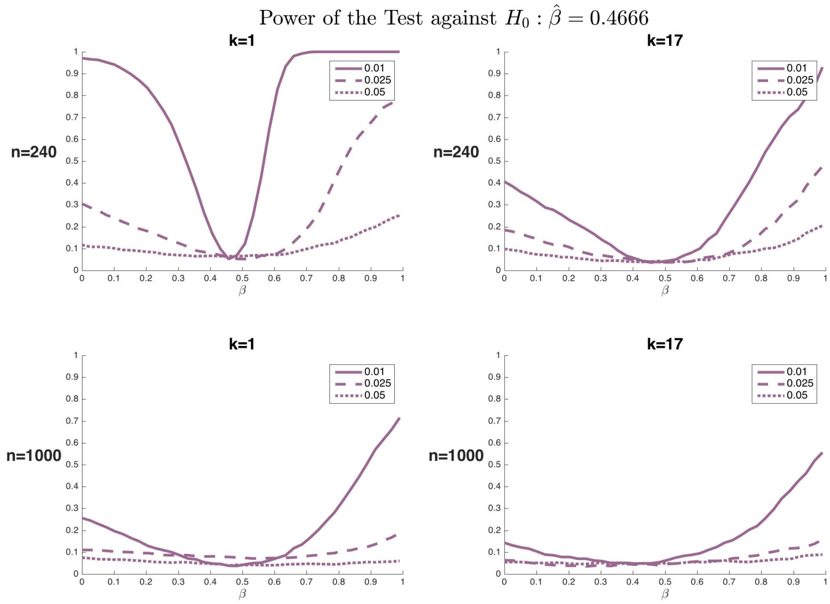


Figure 1.8: Power with different n and k

included terms,

$$\forall k \in \{3, \ldots, \infty\} \quad G^k X = \mu_{0k}\iota + \mu_{1k}X + \mu_{2k}GX + \mu_{3k}GGX + \nu \tag{1.24}$$

The key for recovering a good approximation of Equation 1.23 is that the omitted higher order terms are well approximated by the included terms, in other words, that the R^2 of each Equation 1.24 is high.

Multiplying Equation 1.23 by G gives,

$$Ge \approx G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1}\mathcal{E} \tag{1.25}$$

and by combining Equations 1.23 and 1.25 we can write,

$$e = \beta Ge + \mathcal{E} \tag{1.26}$$

The least squares estimator of β in Equation 1.26 is biased since $Cov(Ge, \mathcal{E}) \neq 0$, but we can explicitly write out the probability limit of the OLS estimator.

$$plim[\hat{\beta}_{OLS}] = \frac{Cov(e, Ge)}{Var(Ge)} \tag{1.27}$$

We evaluate Equation 1.27 by writing out the numerator and denominator and then taking advantage of the fact that $\mathcal{E}$ is the only source of uncertainty and that σ^2 cancels between the numerator and the denominator. This leaves β as the only unknown. After dropping high-order terms in the expansions we solve numerically for β.

$$\begin{aligned} Cov(e, Ge) = Cov\Bigg(&\Big(\mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E}\Big), \\ &\Big(G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1}\mathcal{E}\Big)\Bigg) && (1.28) \\ = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} &\beta^{i+j-1} Cov(G^i\mathcal{E}, G^j\mathcal{E}) && (1.29) \\ = \sum_{k=0}^{\infty} &\beta^k \sum_{l=0}^{k} Cov(G^l\mathcal{E}, G^{k-l+1}\mathcal{E}) && (1.30) \end{aligned}$$

and

$$\begin{aligned} Var(Ge) &= Var(G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1}\mathcal{E}) && (1.31) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \beta^{i+j-2} Cov(G^i\mathcal{E}, G^j\mathcal{E}) && (1.32) \\ &= \sum_{k=0}^{\infty} \beta^k \sum_{l=1}^{k+1} Cov(G^l\mathcal{E}, G^{k-l+2}\mathcal{E}) && (1.33) \end{aligned}$$

Rearranging Equation 1.27 and combining on β, gives us:

$$Cov(e, Ge) - plim[\hat{\beta}_{OLS}]Var(Ge) = 0 \tag{1.34}$$

$$\Rightarrow \left(\sum_{k=0}^{\infty}\beta^k\sum_{l=0}^{k}Cov(G^l\mathcal{E}, G^{k-l+1}\mathcal{E})\right) - \left(plim[\hat{\beta}_{OLS}]\sum_{k=0}^{\infty}\beta^k\sum_{l=1}^{k+1}Cov(G^l\mathcal{E}, G^{k-l+2}\mathcal{E})\right) = 0 \tag{1.35}$$

$$\Rightarrow \sum_{k=0}^{\infty}\beta^k\left(\sum_{l=0}^{k}Cov(G^l\mathcal{E}, G^{k-l+1}\mathcal{E}) - plim[\hat{\beta}_{OLS}]\sum_{l=1}^{k+1}Cov(G^l\mathcal{E}, G^{k-l+2}\mathcal{E})\right) = 0 \tag{1.36}$$

which is a polynomial with respect to β. Since

$$Cov(G^p\mathcal{E}, G^q\mathcal{E}) \approx \frac{\sum diag(G^p * G^q)}{N}\sigma^2_{\mathcal{E}} \tag{1.37}$$

we can approximate the polynomial and solve for its root, which will be our estimate of $\hat{\beta}_{RSEE}$. Since Equation 1.36 is an infinite series, we need to choose an expansion point. Our testing has shown that $k = 3$ works well in practice, so the formula becomes:

$$plim[\hat{\beta}_{OLS}] \approx \beta + \frac{Cov(\mathcal{E}, G\mathcal{E}) + \beta Cov(\mathcal{E}, GG\mathcal{E}) + \beta^2 Cov(\mathcal{E}, GGG\mathcal{E}) + \beta^3 Cov(\mathcal{E}, GGGG\mathcal{E})}{Var(G\mathcal{E}) + \beta^2 Var(GG\mathcal{E}) + 2\beta^2 Cov(G\mathcal{E}, GGG\mathcal{E}) + 2\beta^4 Cov(GG\mathcal{E}, GGGG\mathcal{E})} \tag{1.38}$$

1.4.1 Number of Terms in First-Stage Regression

As previously shown, the goal of the first-stage regression is for the residuals to approximate the series expansion of $\mathcal{E}$. Because we include a small number of terms in Equation 1.21, the residuals might not be a good approximation. However, due to the collinearity between the various powers of the adjacency matrix, there is not much new information in the higher order terms. Figure 1.9 plots the squared correlation coefficient between the residuals and the true series expansion of $\mathcal{E}$ for three different first-stage equations. While only including two terms does worse, especially for low density networks, there is hardly any difference once we move beyond four terms. The first-stage equation is doing a very good job of approximating the error term.

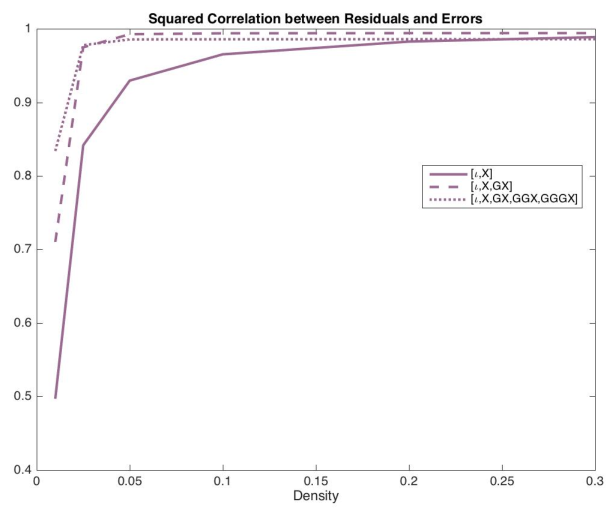


Figure 1.9: Comparison of First-Stage Equations

1.4.2 Small Sample Correction

In a finite sample, the first-stage equation may not well approximate the error term. In particular, this occurs when the sample size is small ($n = 240$) and there are many covariates ($k = 17$). Because of the small sample, one may find a spurious correlation between the error term and the covariates. This causes some of the error term to be projected onto the covariates, and the approximation in the first-stage is not as good. Fortunately, there is an easy fix by including the covariates in the second-stage regression:

$e = \beta Ge + \lambda X + \mu GX + \mathcal{E}$. Although e is uncorrelated with X and GX by construction, in a small sample, Ge is not necessarily uncorrelated with X and GX and including them in the regression improves the estimation. Figure 1.10 shows this improvement when $n = 240$ and also highlights that it is unneccesary for larger samples.

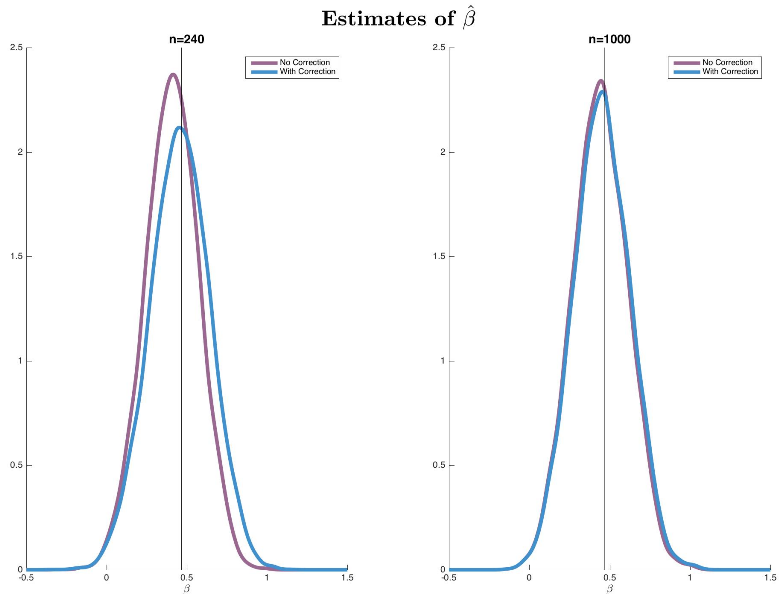


Figure 1.10: Small Sample Correction

1.4.3 Standard Error Estimation

Standard errors can be estimated either via the delta method or by the bootstrap. Notice from Equation 1.38 that we solve $\hat{\beta}_{OLS} = f(\beta)$, for some implicit function f. Therefore $\hat{\beta}_{RSEE} = f^{-1}(\hat{\beta}_{OLS})$ and we can apply the delta method:

$$\sqrt{n}\left(f^{-1}(\hat{\beta}_{OLS}) - f^{-1}(\beta)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\hat{\beta}_{OLS}}[f^{-1}]'(\beta)^2\right)$$
$$\Rightarrow \sqrt{n}\left(\hat{\beta}_{RSEE} - \beta\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2_{\hat{\beta}_{OLS}}}{f'(\beta)^2}\right)$$

Alternatively, we generate bootstrapped standard errors. For an estimated $\hat{\beta}_{RSEE}$, we construct $Y - \hat{\beta}_{RSEE}GY$ and regress this on $[\iota, X, GX]$. Since this is no longer a simultaneous system, we can get estimates of α, γ, and δ via OLS.

$$Y - \hat{\beta}_{RSEE}GY = \alpha\iota + (\beta - \hat{\beta}_{RSEE})GY + \gamma X + \delta GX + \mathcal{E} \tag{1.39}$$

From these estimates, we generate residuals.

$$J = \left(Y - \hat{\beta}_{RSEE}GY\right) - \left(\hat{\alpha}\iota + \hat{\gamma}X + \hat{\delta}GX\right) \tag{1.40}$$

We then multiply the residuals by either -1 or 1 and reassign them to a new observation (with replacement). With the new residuals, we compute a new Y_b and then estimate a new $\hat{\beta}$.

$$p_i \in \{-1, 1\} \tag{1.41}$$
$$\mathcal{E}_j^b = J_i p_i \tag{1.42}$$
$$Y^b = (I_n - \hat{\beta}_{RSEE}G)^{-1}(\hat{\alpha}\iota + \hat{\gamma}X + \hat{\delta}GX + \mathcal{E}^b) \tag{1.43}$$
$$\Rightarrow \hat{\beta}_{RSEE}^b \tag{1.44}$$

Doing this a number of times gives a distribution of $\hat{\beta}$, and to construct the hypothesis test, we take the 2.5th and 97.5th percentiles of the distribution and see if the true β is contained in that region.

1.5 Simulation Results

We use the same simulation conditions as in Section 3 and estimate $\hat{\beta}$ using both 2SLS and RSEE. Figure 1.11 plots the distribution of the estimated $\hat{\beta}$s.

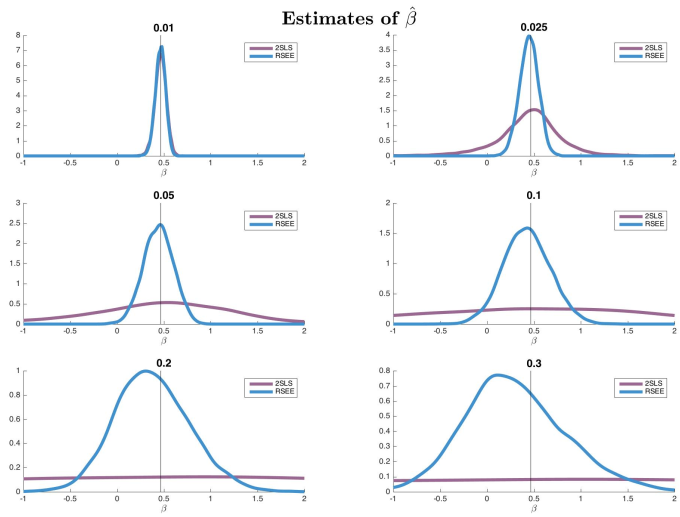


Figure 1.11: Distribution of $\hat{\beta}$

The RSEE estimates are very similar to 2SLS for very sparse networks and are significantly more precise for more dense networks. Figures 1.12 and 1.13 show the size and power, respectively. While 2SLS tests are undersized, our alternative tests are somewhat oversized. This is less so for the bootstrap tests, which have close to nominal size for moderately dense networks. For power, RSEE does much better, especially for more dense networks. For example, for networks with a 10 percent density the standard 2SLS tests are essentially uninformative against an alternative of 0, with size equal 0.021 and power equal 0.029. The bootstrap version of our new estimator has empirical size and power of 0.081 and 0.67, respectively.

It is often useful to consider size-adjusted power when evaluating a proposed estimator. To do this, we calculate the 2.5th and 97.5th percentiles of the empirical t-

distribution. These correspond to the empirical critical values for that test. We then use those critical values when computing the power. Figure 1.14 plots the size-adjusted power. On a size-adjusted power, the bootstrap inference is much preferred to 2SLS results.

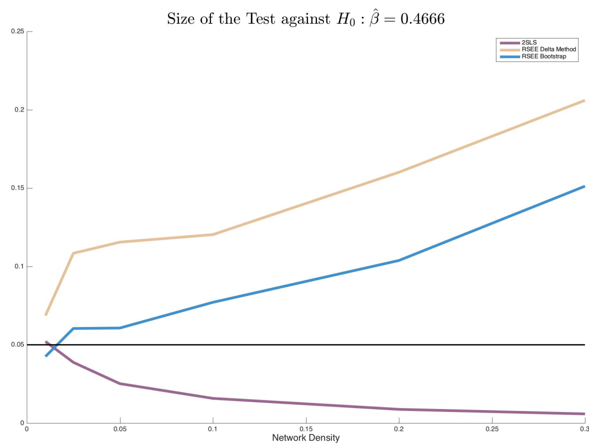
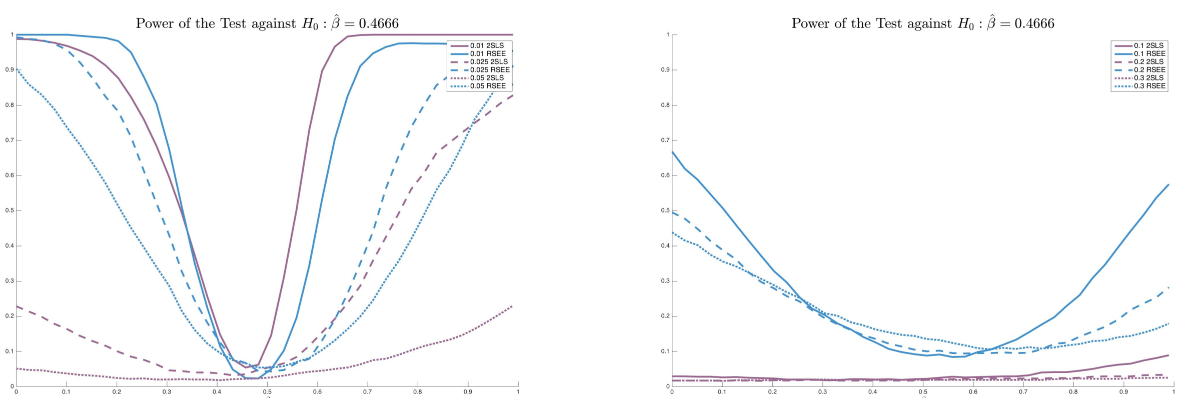


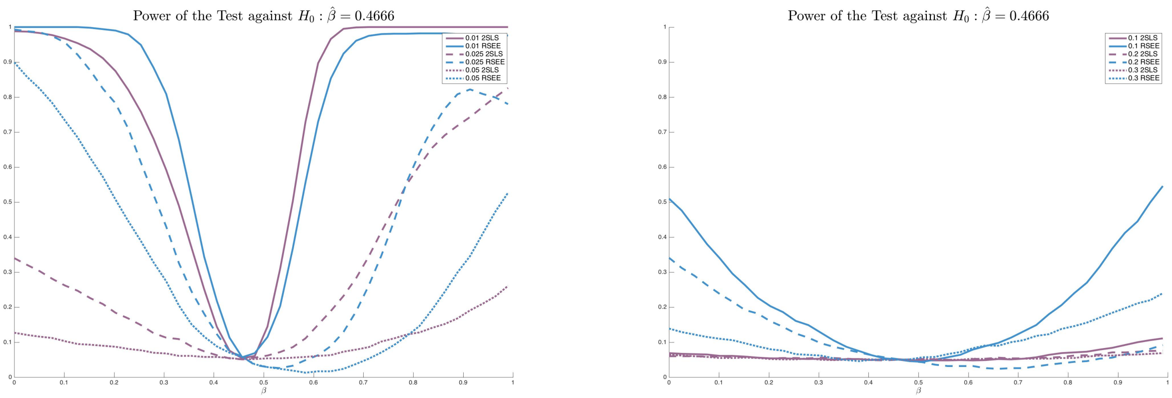
Figure 1.12: Size



(a) Low Density (b) High Density

Figure 1.13: Power

While it is clear that the RSEE has much more power, we might be concerned that we gain this power at the expense of more bias. Figure 1.15 shows the comparison of the median bias.



(a) Low Density (b) High Density

Figure 1.14: Size-Adjusted Power

RSEE is quite biased for very dense networks, however, the bias of RSEE is comparable to 2SLS for reasonably sparse networks.

1.5.1 Standard Error Estimation

The estimated standard errors from the alternative method are too small when we use the delta method. This is likely because this is for a second stage with generated regressors, since Equation 1.27 is a function of the residuals, e, which are a product of Equation 1.21. To test this, we plug in the true (series expansion of) error terms into Equation 1.27 and re-run the delta method. Figure 1.16 plots the size of both the original delta method and that without the generated regressors. This is evidence that the delta method is not providing accurate standard errors because of a generated regressors problem.

On the other hand, the bootstrap seems to be providing much more reasonable estimates of the standard error. Figure 1.17 shows the distribution of the bootstrapped standard errors relative to the empirical standard deviation. Here we can see the bootstrap is doing a reasonable job of approximating the standard error.

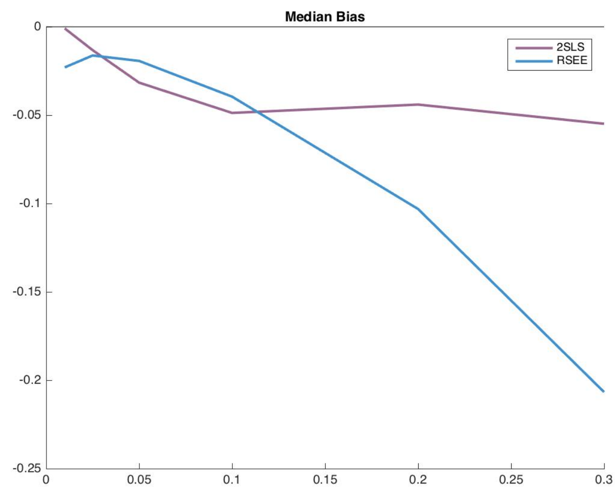


Figure 1.15: Median Bias

1.5.2 Different n and k

We also compare our estimator across different n and k as in Section 1.3.3. Figure 1.18 shows the (size-adjusted) power of both 2SLS and RSEE. While 2SLS gets much worse in all cases, the power of RSEE is essentially unchanged, suggesting that it performs well in many settings. Figure 1.19 shows that the power of RSEE is not gained at the expense of bias. It also shows that the 2SLS estimates for many covariates are potentially quite biased.

1.6 Empirical Application

To illustrate our method, we replicate the empirical results from BDF and then estimate peer effects for each school in the BDF sample. Following BDF, we use the In-School sample from Wave I of the National Longitudinal Study of Adolescent to Adult Health (Add Health). This is a nationally representative sample of 80 high schools and 52 middle schools. Each student in the sample was asked to fill out a questionnaire

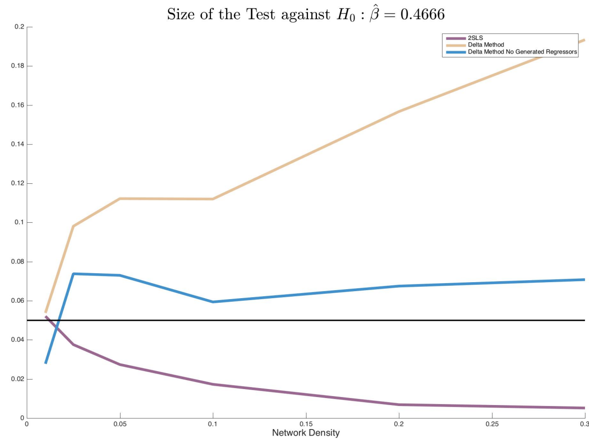


Figure 1.16: Comparison of Delta Methods

with questions pertaining to social and demographic characteristics, school activities and behavior, and parent education and occupation. The questionnaire also asked each student to list up to five male and five female friends. This friendship elicitation is used to construct the network.[5] The dependent variable is an index of participation in recreational activities and the covariates include characteristics of the student as well as their parents. We present a replication of the estimates from BDF, differing in that we have a slightly different sample. Table 1.1 shows that the summary statistics in BDF and our sample are quite similar.

We follow BDF and construct one large block diagonal friendship network where each individual school network is placed on the diagonal. By combining all of the schools we have constructed a very sparse network: in our sample the density is $5.67 * 10^{-5}$. We then estimate the model using 2SLS and RSEE. Table 1.2 compares the results between BDF and our sample. Once again, there are some differences between the two samples, with the most important being that the estimated peer effect in our sample is larger by a fraction of a standard error. Most of the other estimated coefficients are also fairly similar. The

[5] While the limit of ten total friends means that some friendships might be censored, in practice very few students list all ten.

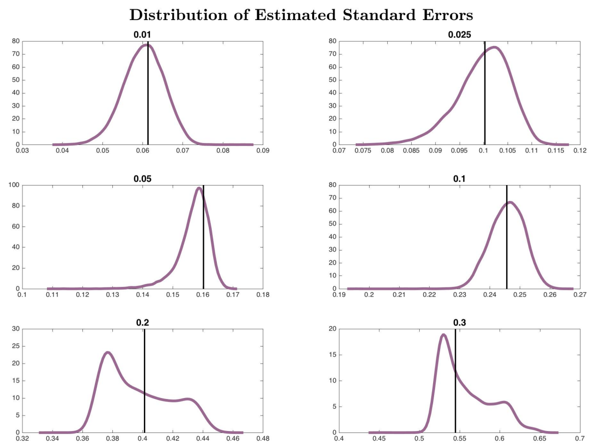


Figure 1.17: Distribution of Standard Errors

third column of results estimates the model using RSEE. The RSEE coefficient estimate is within one standard error of the 2SLS estimate. The estimated standard error on the peer effect is smaller by an order of magnitude when using RSEE.

To highlight the benefits of our method for less dense networks, we estimate the peer effects for each school. Figure 1.20 shows a histogram of the individual school network densities. Most of the schools are around the 1% density level, but all of them are significantly more dense than the combined network by a factor of nearly 1,000. For each school we estimate peer effects using both 2SLS and RSEE.[6]

Figure 1.21 plots the kernel density of the peer effect coefficients. The RSEE estimates are nicely distributed around 0.3, with an interquartile range of 0.22, while the 2SLS estimates are quite disperse, with an interquartile range of 1.21. Figure 1.22 plots the kernel density of the peer effect standard errors. The RSEE estimates are much more precise, with a median standard error of 0.088, while the corresponding median 2SLS standard error is 0.89. This means that for any reasonable estimate of the peer effect (between -1 and 1), 2SLS is unlikely to be able to reject the null hypothesis of no

[6]Some of the schools are completely homogeneous in one of the included covariates (Race, Born in US, etc.) so for those schools that covariate is dropped.

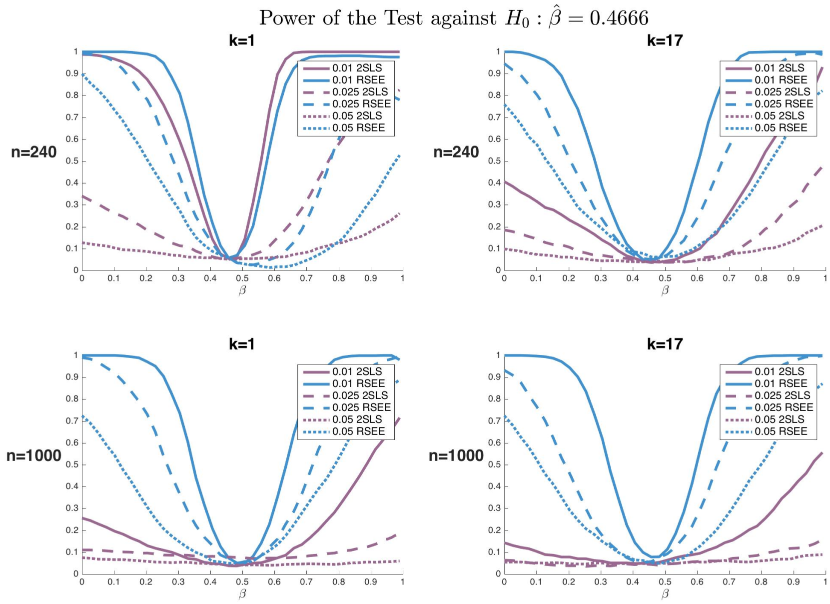


Figure 1.18: Power with different n and k

peer effects. In other words, 2SLS will be uninformative about peer effects in individual schools.

Our method will allow for the estimation of individual school peer effects. This is important when the unit of interest is an individual school and will allow for testing for heterogeneity between schools. A Wald test of joint equality between the individual peer effects and the full network peer effect (0.365) is strongly rejected ($\chi^2 = 488$), which suggests there is substantial heterogeneity in peer effects between the schools.

1.7 Conclusion

In this paper we replicate the findings of Bramoullé, Djebbari and Fortin (2009) and highlight the weak instrument problem that results in limited power for denser networks. We propose an alternative estimation technique that does not rely on a weak instrument and show it performs much better than 2SLS for dense networks.

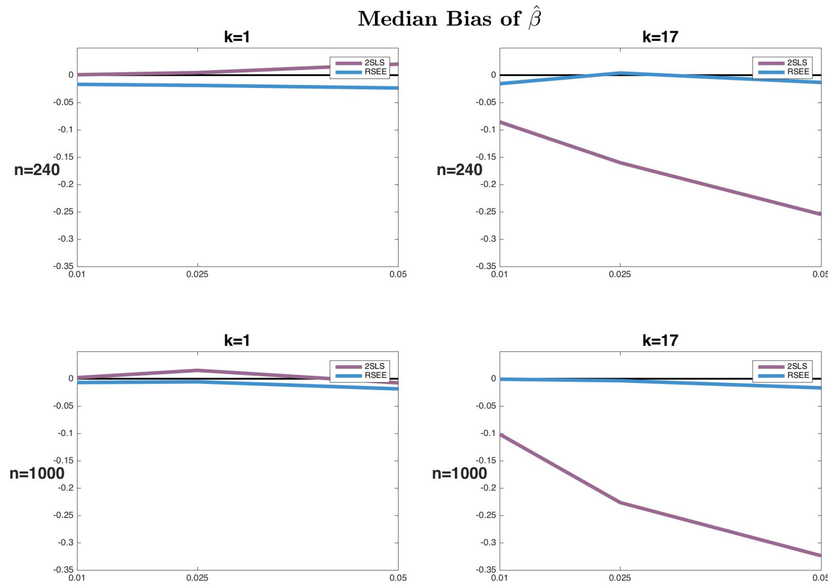


Figure 1.19: Median Bias with different n and k

1.8 Permissions and Attributions

Chapter 1 is a result of collaboration with Richard Startz.[7]

[7] We thank Doug Steigerwald and members of the UCSB Econometrics Working Group for helpful comments and suggestions. We acknowledge support from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1121053) and NSF CNS-0960316. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

	BDF		Our Sample	
Variable	Mean	SD	Mean	SD
Recreational Activities	2.122	1.267	1.925	1.440
Age	14.963	1.682	14.965	1.683
Female	0.535	0.499	0.535	0.499
Race is white only	0.619	0.486	0.613	0.487
Born in the US	0.928	0.259	0.923	0.266
Mother Present	0.929	0.257	0.930	0.255
Father Present	0.779	0.415	0.776	0.417
Grade 6 to 8	0.263	0.440	0.264	0.441
Grade 9 or 10	0.406	0.491	0.408	0.491
Grade 11 or 12	0.331	0.471	0.329	0.470
Parents' labor force participation	0.965	0.184	0.907	0.290
Mother No HS	0.097	0.296	0.102	0.302
Mother is HS grad	0.284	0.451	0.320	0.467
Mother more than HS but no college	0.276	0.447	0.157	0.363
Mother College grad	0.206	0.404	0.279	0.448
Mother went to school but unknown level	0.066	0.248	0.072	0.259
Father No HS	0.081	0.273	0.084	0.277
Father is HS grad	0.211	0.408	0.231	0.421
Father more than HS but no college	0.240	0.427	0.118	0.323
Father College grad	0.178	0.383	0.270	0.444
Father went to school but unknown level	0.069	0.253	0.074	0.261
Number of Observations	55208		53601	

Table 1.1: Summary Statistics

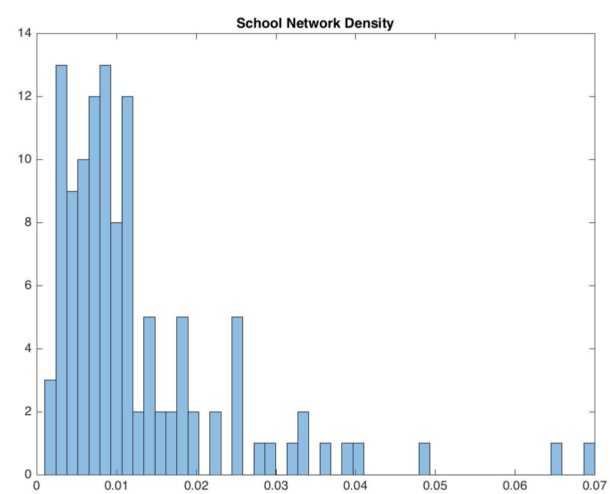


Figure 1.20: Histogram of School Network Density

		BDF		Our Sample			
		2SLS		2SLS		RSEE	
	Variable	Coef	SE	Coef	SE	Coef	SE
	Age	−0.022	0.011	0.086	0.004	−0.049	0.007
	Female	0.213	0.015	0.219	0.015	0.185	0.012
	Race is white only	−0.106	0.020	−0.090	0.019	−0.085	0.014
	Born in the US	−0.052	0.033	0.021	0.029	0.000	0.023
	Mother Present	−0.013	0.036	−0.051	0.036	−0.091	0.028
	Father Present	−0.018	0.029	−0.020	0.029	−0.040	0.023
	Grade 9 or 10	0.011	0.096	−0.125	0.043	−0.006	0.039
	Grade 11 or 12	0.021	0.102	−0.328	0.048	−0.029	0.046
Own Characteristics	Mother is HS grad	−0.005	0.027	0.038	0.027	0.068	0.020
	Father is HS grad	0.047	0.029	0.066	0.028	0.092	0.022
	Mother more than HS but no college	0.146	0.029	0.177	0.031	0.221	0.023
	Father more than HS but no college	0.167	0.030	0.158	0.033	0.236	0.025
	Mother College grad	0.137	0.033	0.271	0.032	0.342	0.022
	Father College grad	0.127	0.031	0.256	0.031	0.327	0.023
	Mother went to school but unknown level	−0.010	0.038	−0.016	0.037	−0.010	0.028
	Father went to school but unknown level	−0.067	0.038	−0.029	0.038	−0.051	0.029
	Parents' labor force participation	0.083	0.040	0.124	0.029	0.082	0.022
	Age	−0.061	0.020	−0.064	0.018	−0.060	0.004
	Female	0.008	0.048	−0.035	0.048	−0.055	0.020
	Race is white only	−0.019	0.045	−0.035	0.042	0.080	0.020
	Born in the US	0.042	0.066	−0.014	0.061	0.106	0.032
	Mother Present	0.107	0.064	0.105	0.081	0.119	0.050
	Father Present	−0.109	0.053	−0.122	0.064	−0.133	0.040
	Grade 9 or 10	−0.034	0.186	0.011	0.094	0.040	0.043
	Grade 11 or 12	0.100	0.194	0.159	0.125	0.256	0.049
Friends' Characteristics	Mother is HS grad	−0.047	0.050	−0.017	0.059	0.113	0.036
	Father is HS grad	0.172	0.055	0.113	0.062	0.186	0.040
	Mother more than HS but no college	−0.038	0.068	0.073	0.074	0.197	0.041
	Father more than HS but no college	0.091	0.067	0.021	0.084	0.248	0.045
	Mother College grad	−0.031	0.081	−0.024	0.098	0.214	0.039
	Father College grad	0.124	0.061	0.118	0.083	0.314	0.041
	Mother went to school but unknown level	−0.094	0.070	−0.053	0.083	−0.027	0.052
	Father went to school but unknown level	0.150	0.075	0.248	0.085	0.191	0.054
	Parents' labor force participation	0.151	0.073	0.085	0.065	0.013	0.040
Peer Effect	Endogenous Effect	0.467	0.256	0.492	0.195	0.365	0.006
	Number of Observations	55208		53601		53601	

Table 1.2: Full Sample Results

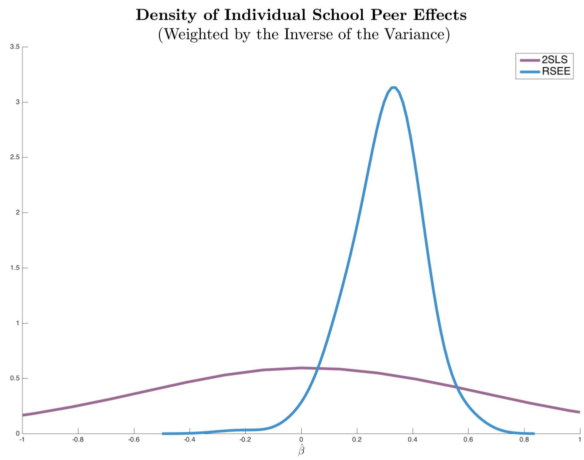


Figure 1.21: Kernel Density of Peer Effect Coefficients

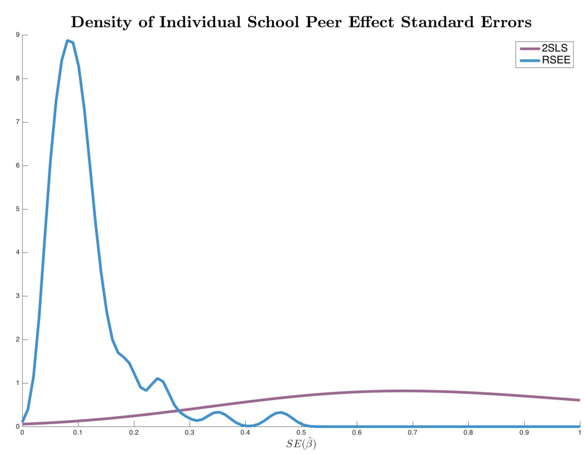


Figure 1.22: Kernel Density of Peer Effect Standard Errors

Chapter 2

Do Employers Learn from Public, Subjective, Performance Reviews?

2.1 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 unwise 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, there are a number of reasons why this feedback might not provide useful information. Reviews might be biased because of a fear of retaliation, lack of anonymity, strategic behavior, or sorting into reviewing (Tadelis, 2016). Further, in online labor markets, since 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 there is bias in the reviews, 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 similar testable predictions to the symmetric employer learning literature (Altonji and Pierret, 2001). It also generates additional hypotheses about asymmetric learning from individual reviews. Using data from a large online labor market, I find strong evidence that public performance reviews do provide information to the market. First, I show that aggregate descriptive statistics provide support for the symmetric employer learning hypotheses about the variance of wages and the correlation between wages and worker characteristics. Then I focus specifically on learning that occurs across firms and 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

[1] I will use employer and firm synonymously throughout.

evidence suggests that, despite an incentive to bias reviews, employers are leaving public performance reviews that do provide information about workers' productivity. I then test a mechanism for this result. Using data on applications, I show that workers are unwilling to renew employment contracts with firms that left them a negative review. This will incentivize firms who wish to rehire a worker to leave an informative review. Despite this mechanism, 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 rehiring considerations help overcome the incentive for employers to bias their reviews.

This paper contributes to the growing literature on reputation systems.[2] A theme of this literature is the documentation of various biases in reviews. A big concern is a fear of retaliation, where one side worries that if they leave a bad (and honest) review, the other side will retaliate in some form. This has been documented in eBay (Bolton, Greiner and Ockenfels, 2012) and Airbnb (Fradkin et al., 2016). This can also lead to significant review inflation, where nobody leaves bad reviews, as seen in oDesk (Horton and Golden, 2015) and Airbnb (Zervas, Proserpio and Byers, 2015). Another, related, concern is sorting into reviews: if the only people who leave reviews are those who are really happy or really upset, we will get an incorrect picture of the market. This phenomena is documented on Ebay (Dellarocas and Wood, 2007; Nosko and Tadelis, 2015; Masterov, Mayer and Tadelis, 2015) and Amazon (Hu, Zhang and Pavlou, 2009). There are additional concerns about users purchasing reviews and/or leaving fake reviews. Users may want to boost their profile by acquiring additional feedback through improper

[2]See Tadelis (2016) and Luca (2016) for recents surveys.

channels. There is evidence of significant review buying on eBay (Brown and Morgan, 2006) and Taobao (Xu et al., 2015), as well as fake reviews on Yelp (Luca and Zervas, 2015) and Tripadvisor (Mayzlin, Dover and Chevalier, 2014). Finally, there is a potential re-hiring bias in online labor markets. 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 hide their talented workers by not providing public information about their skill (for example by promoting them). All of these biases limit the functionality and usefulness of reputation systems.

There has also been some related work on online labor markets. Benson et al. (2015) run a field experiment showing that the employer's reputation affects their ability to hire workers. On the worker side, Pallais (2013) hires workers in an online labor market and finds that providing jobs to inexperienced workers significantly improves their later outcomes. Similarly, Agrawal, Lacetera and Lyons (2016) show that work experience is very important in these markets, especially for workers from developing countries. Both of these studies are similar to mine, but focus on experience rather than reviews. While it is true that the two are correlated, we should expect reviews to be a better signal of worker ability. However, given the large literature on potential biases, it is unclear how useful reviews are in these markets. Mill (2011) looks at reviews in his study of an online labor market, however he uses reviews as a measure of weighted experience by counting the number of different types of reviews. The goal of the current paper is to explicitly measure the effect of receiving a single review of different types and how employers learn from that review.

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 proposed that all 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 digitized.

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. Sections 4 and 5 test these hypotheses. Section 6 investigates the effect of reviews on worker attrition. Section 7 discusses and tests a possible mechanism for these results and Section 8 investigates other possible strategic behavior by firms.

2.2 Data

2.2.1 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, 2015). 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.[3] The employer grades the worker out of five stars in six different categories: Availability, Communication,

[3]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.

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.

2.2.2 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.[4] 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 to focus on the early careers of workers. Table 2.1 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.[5] 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.[6] 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,

[4]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.

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

[6]However, the hourly wage is comparable to the average hourly wage in India and the Philippines.

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.

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

$$\text{Review Type} = \begin{cases} \text{Good} & \text{if Score} = 5 \\ \text{Bad} & \text{if } 4 \leq \text{Score} < 5 \\ \text{Really Bad} & \text{if Score} < 4 \\ \text{No Review} & \text{if no review was left} \end{cases} \tag{2.1}$$

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

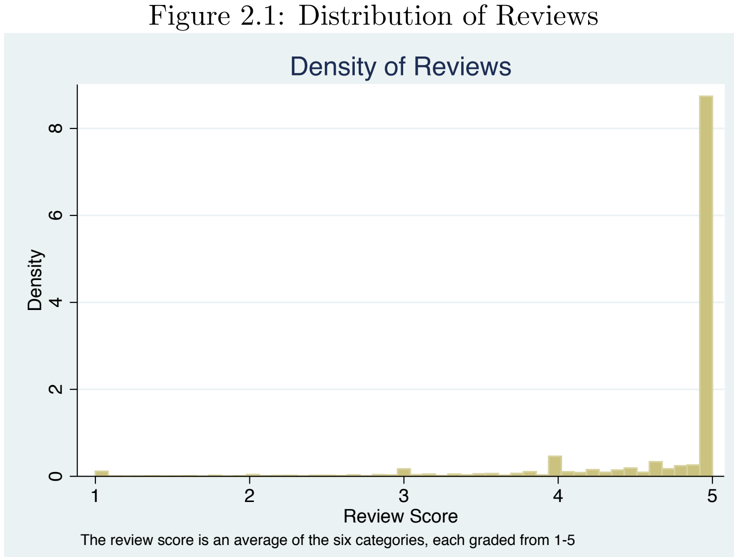


Table 2.1: Summary Statistics for Workers

High Income Country: OECD	0.133
High Income Country: Non-OECD	0.00964
Upper Middle Income Country	0.0539
Lower Middle Income Country	0.659
Low Income Country	0.144
Bachelor's Degree	0.498
Master's Degree	0.0751
Number of Outside Experiences	2.695
Number of oDesk Exams	2.131
Average Wage ($)	5.502
Average Job Duration (Hours)	170.3
Average Job Duration (Days)	29.01
Average Career Length (Jobs)	6.736
Average Number Jobs at One Time Conditional on Having a Job	1.250
N	16189

2.3 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): $\theta_i = \gamma x_i + \nu_i$.[7] 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 observed signal of productivity: $r_{it} = g(y_{it})$ for some weakly monotonically increasing function g with at least one strict increase. Thus, honest revelation ($r_{it} = y_{it}$), shading by a constant, known factor ($r_{it} = 0.7 * y_{it}$), and a threshold model $\left(r_{it} = \begin{cases} 5 & \text{if } y_{it} \geq \bar{y} \\ 1 & \text{if } y_{it} < \bar{y} \end{cases} \right)$ 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

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

contain no information about the worker and therefore have no effect of wages:

$$w_{ijt+1} = \mathbb{E}[\theta_i|x_i] + \eta_{ijt+1} \tag{2.2}$$

This is the extreme case where no one believes the reviews left by other firms.[8]

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

$$w_{it+1} = \mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it}] + \eta_{ijt+1} \tag{2.3}$$

An informative world is one in which at least some employers are providing their private information to the market.[9] 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 or uninformative reviews. Finally, a worker may not want to be rehired by an employer that left a poor or uninformative review. All of these factors potentially enter into the employer's decision process.

This simple model is able to generate the three main results from the symmetric employer learning literature (Altonji and Pierret, 2001). 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 cross-sectional variance of wages increases with t. In an uninformative world, the variance is constant.*

When a worker is first hired, a firm does not know much about his productivity, so,

[8] This would also be the case if firms did not leave reviews at all.

[9] A more formal treatment of this model is presented in Appendix A.2.

conditional on observables, all workers should be paid similarly. In an informative world, firms observe the worker's reviews and their wage will reflect their productivity. Highly productive workers will have more wage growth than lower productivity workers, and thus the variance of overall wages will increase over time. In an uninformative world, reviews provide no information about the worker, so they have no effect on wages.

The other two results describe the relationship between wages and worker characteristics.

Proposition 2. *In an informative world, the marginal effect of the observable characteristic on the conditional mean of wages is strictly positive and decreasing with t. In an uninformative world, the marginal effect is strictly positive and constant.*

Proposition 3. *In an informative world, the marginal effect of the worker's average review score on the conditional mean of wages is strictly positive and increasing with t. In an uninformative world, the marginal effect is zero.*

In an informative world, 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 their reviews, the effect of the average review on wages will increase over time.[10] In an uninformative world, there is no information in reviews, so firms are unable to update their beliefs about workers.

This theoretical framework also allows for more direct tests of asymmetric learning. By isolating the effect of the most recent review on wages, I can focus on workers who

[10] Note that this is slightly different from the symmetric employer learning models. In those models there is some measure of productivity that is unobservable to the firm (AFQT score). In this model, firms are able to observe the reviews, and it's the fact that the average review gets more precise over time that generates the predictions.

were never previously employed by the hiring firm. This means the hiring firm does not have any private information about the worker and therefore must rely on the public reviews. Thus, I can test whether other firms are learning about worker productivity through individual reviews. To do this, I first difference Equations (2.2) and (2.3). This will difference out the effect of previous reviews.[11] In an uninformative world:

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i|x_i] + \eta_{ijt+1} - (\mathbb{E}[\theta_i|x_i] + \eta_{ikt}) \tag{2.4}$$

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:

$$\Delta w_{ijt+1} = w_{ijt+1} - w_{ikt} = \mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it-1}, r_{it}] + \eta_{ijt+1} - (\mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it-1}] + \eta_{ikt}) \tag{2.5}$$

These two wage equations yield the following two hypotheses about the relationship between wage changes and worker characteristics.

Proposition 4. *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 ad-

[11]The previous reviews will not completely difference out because the weighting of each previous review in the conditional expectation will likely change with the addition of another review. For example, in the normal learning models, each review would have equal weights. However, comparing wages at two time periods, there will be different numbers of reviews in the conditional expectations and thus different weights, so the effect of any given review will not completely difference out. However, under reasonable assumptions about the conditional expectation, the contribution of the most recent review will be much greater than that of any of the previous reviews (since it will only be in the later conditional expectation).

ditional review is lowered. In an uninformative world, reviews provide no information about the worker, so they have no effect on wages.

Proposition 5. *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 identical workers with no reviews except that one has more education than the other. The more educated worker will receive a higher wage in the first period. In an informative world, conditional on receiving the same review, the worker with the lower education will see a larger increase (or a smaller decrease) in their wage than the worker with more education. This effect will taper off over time as more of the weight is placed on the reviews and less on education.

2.4 Aggregate Evidence

To motivate the role of employer learning in this market, I present evidence for the first three propositions. Figure 2.2 plots the cross-sectional variance of wages for the first 20 jobs of a worker's career.[12] There are clear increases in the variance of wages over time ($p < 0.01$ for the slope coefficient) which provides support for Proposition 1.

To test Propositions 2 and 3 I regress (log) wages on a number of worker characteristics, including their country and their average review score. I estimate the following

[12] To eliminate the effect of attrition from oDesk on the variance, only workers who stay for at least 20 jobs are included.

Figure 2.2: Variance of Wages

equation:

$$\log(w_{it}) = \alpha + \beta HighIncome_i + \gamma AvgScore_{it} + \delta g(Exp_{it}) \\ + \theta HighIncome_i * g(Exp_{it}) + \kappa AvgScore * g(Exp_{it}) + \boldsymbol{\lambda X_{it}} + \epsilon_{it} \tag{2.6}$$

where $HighIncome_i$ 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 $\boldsymbol{X_{it}}$ are other worker characteristics such as education and exam scores.[13] 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 include the country of the worker as the easy to observe variable. Following Altonji and Pierret (2001), I model experience as a

[13]For 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.

cubic polynomial and look at the relationship between worker characteristics and wages over time.[14] Figure 2.3 plots the marginal effect of the easily observable characteristic (whether the worker is from a high income country) and the effect of the average review.[15] The importance of the worker's country in determining wages decreases over time, while the effect of the average review score increases, which provides support for Propositions 2 and 3.

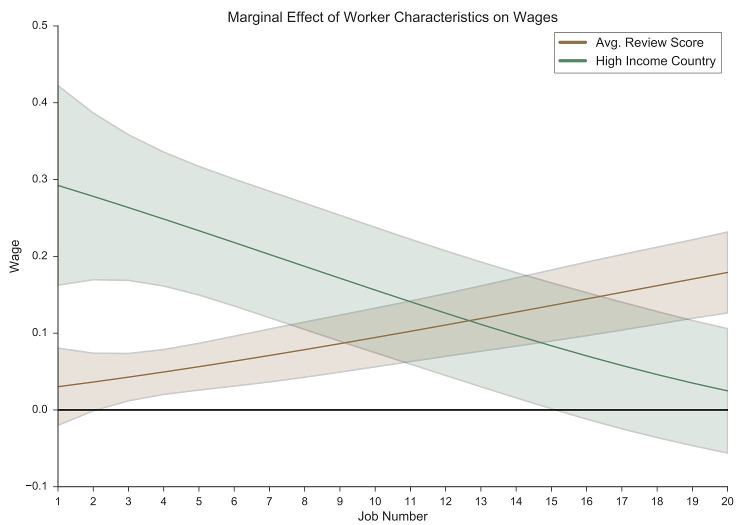


Figure 2.3: Marginal Effects of Worker Characteristics on Wages

These results suggest that the public performance reviews are informative and the labor market learns about the ability of workers over time. While these results are consistent with a model of symmetric learning, the theoretical framework allows for predictions about asymmetric learning from reviews. To test these predictions, I need to directly measure the effect of a single review on how much workers are paid by other employers.

[14]Again, I limit my sample to workers who stay for at least 20 jobs.
[15]Appendix A.1 provides the regression estimates.

2.5 Learning by Other Firms

To focus on across-firm learning, I now limit my sample to the 80% of jobs in which the worker was never previously employed by the hiring firm. 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.[16] I estimate the following equation:

$$\begin{aligned}\Delta \log(w_{ijt}) &= \boldsymbol{\beta} \boldsymbol{ReviewType_{it}} + \gamma After_{it} + \boldsymbol{\delta}(\boldsymbol{ReviewType_{it}} * After_{it}) \\ &+ \boldsymbol{\rho} \boldsymbol{CountryIncome_i} + \boldsymbol{\tau}(\boldsymbol{CountryIncome_i} * After_{it}) \\ &+ \boldsymbol{\lambda X_{it}} + \boldsymbol{\mu J_{it}} + \nu_i + \kappa_j + \epsilon_{ijt} \qquad \mathbb{E}[\epsilon_{ijt}|\boldsymbol{Z_{ijt}}] = 0\end{aligned} \tag{2.7}$$

where $\boldsymbol{ReviewType_{it}}$ are indicators for the type of the most recent review and $\boldsymbol{CountryIncome_i}$ are indicators for the income classification of the worker's country. $\boldsymbol{X_{it}}$ are worker characteristics, such as education, exam scores, and previous review scores and $\boldsymbol{J_{it}}$ are job characteristics, which include indicators for different words that show up in the job titles.[17] ν_i are worker fixed effects and κ_j are firm fixed effects, which are

[16] The results are consistent for a number of different cut-points and non-parametric experience. See Appendix A.8.

[17] 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.

included depending on the specification. Finally, $\boldsymbol{Z_{ijt}}$ contains all the included variables.

Table 2.2 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 4: 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 5. 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. After only five jobs firms are still willing to pay a premium for workers from higher income countries. While the decrease is 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 (go from a 4% wage increase to only a 2% wage increase) 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 (go from a 11% wage decrease to only a 3.5% decrease).

We can further investigate these learning results by splitting the workers and firms into different categories. First, we might think that there are differences in how firms

Table 2.2: Effect of Reviews on Wage Changes

		(1)	(2)	(3)	(4)
Really Bad Review	First 5 Jobs	-0.0291	-0.0350	-0.0593*	-0.108**
		(0.0177)	(0.0229)	(0.0281)	(0.0387)
	After 5 Jobs	-0.00142	0.00662	-0.0152	-0.0320
		(0.00815)	(0.00985)	(0.0155)	(0.0197)
No Review	First 5 Jobs	0.0235	0.0265	0.0381	0.0112
		(0.0146)	(0.0196)	(0.0237)	(0.0314)
	After 5 Jobs	0.0119^{+}	0.0117	0.0178	0.0223
		(0.00614)	(0.00758)	(0.0123)	(0.0157)
Good Review	First 5 Jobs	0.0202*	0.0192	0.0507**	0.0404^{+}
		(0.00958)	(0.0127)	(0.0155)	(0.0222)
	After 5 Jobs	0.0134**	0.00950	0.00712	0.00205
		(0.00514)	(0.00630)	(0.0103)	(0.0130)
High Income Country	First 5 Jobs	-0.0434**		-0.109**	
		(0.0113)		(0.0238)	
	After 5 Jobs	-0.0249**		-0.0897**	
		(0.00645)		(0.0182)	
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

learn about worker's from different countries. In particular, since most firms are from high income countries, we might expect them to learn about workers from high income countries differently from workers from low income countries. Table 2.3 splits workers into two categories according to the World Bank classification of their country. Here we can see that the learning results are primarily driven by the workers from low income countries. Reviews do not have a significant effect on workers from high income countries, but they are quite important for workers in developing countries. These results are consistent with Agrawal, Lacetera and Lyons (2016) but provide evidence that the effect is going through reviews and not just work experience.

Table 2.3: Effect of Reviews on Wage Changes by Worker Country

			(1)	(2)	(3)	(4)
Really Bad Review	**Low Income**	First 5 Jobs	-0.0340^{+}	-0.0395^{+}	-0.0708*	-0.110**
			(0.0184)	(0.0239)	(0.0292)	(0.0401)
		After 5 Jobs	-0.000891	0.00581	-0.0138	-0.0312
			(0.00866)	(0.0104)	(0.0163)	(0.0202)
	High Income	First 5 Jobs	0.0123	0.0192	0.0509	-0.0512
			(0.0477)	(0.0631)	(0.0674)	(0.106)
		After 5 Jobs	-0.00313	0.0151	-0.0257	-0.0455
			(0.0205)	(0.0272)	(0.0457)	(0.0711)
Good Review	**Low Income**	First 5 Jobs	0.0228*	0.0229^{+}	0.0541**	0.0500*
			(0.0104)	(0.0136)	(0.0165)	(0.0231)
		After 5 Jobs	0.0140*	0.0101	0.00800	0.00320
			(0.00544)	(0.00661)	(0.0108)	(0.0135)
	High Income	First 5 Jobs	0.00647	-0.00486	0.0231	-0.0736
			(0.0195)	(0.0279)	(0.0403)	(0.0699)
		After 5 Jobs	0.0106	0.00437	-0.00160	-0.0118
			(0.0150)	(0.0200)	(0.0319)	(0.0477)
Worker Fixed Effects			No	Yes	No	Yes
Firm Fixed Effects			No	No	Yes	Yes
R^2			0.0450	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. Workers are classified by the World Bank category of their country. 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

We can also split the sample by looking at the work experience of firms. We might expect that there is learning going on by firms about how to use reviews to evaluate workers. Table 2.4 splits firms into two experience categories by the number of jobs they have completed. More experienced firms are much more likely to use reviews to learn about worker productivity. Almost all of the learning results from Table 2.2 are coming from experienced firms. This suggests there is a learning curve in this online platforms.

Table 2.4: Effect of Reviews on Wage Changes by Firm Experience

			(1)	(2)	(3)	(4)
Really Bad Review	**Low Exp Firm**	First 5 Jobs	0.0312	0.00688	-0.0326	-0.0344
			(0.0270)	(0.0356)	(0.0503)	(0.0751)
		After 5 Jobs	0.000950	0.0122	-0.00700	-0.00928
			(0.0135)	(0.0156)	(0.0355)	(0.0451)
	High Exp Firm	First 5 Jobs	-0.0532*	-0.0520+	-0.0723*	-0.118**
			(0.0213)	(0.0268)	(0.0299)	(0.0398)
		After 5 Jobs	-0.00154	0.00873	-0.0122	-0.0288
			(0.0104)	(0.0123)	(0.0165)	(0.0205)
Good Review	**Low Exp Firm**	First 5 Jobs	0.00352	-0.00707	0.00356	-0.0135
			(0.0154)	(0.0201)	(0.0310)	(0.0481)
		After 5 Jobs	0.0192*	0.0194+	0.0219	0.0312
			(0.00896)	(0.0104)	(0.0227)	(0.0280)
	High Exp Firm	First 5 Jobs	0.0299*	0.0353*	0.0528**	0.0442+
			(0.0121)	(0.0155)	(0.0172)	(0.0235)
		After 5 Jobs	0.0111	0.0116	0.0123	0.00493
			(0.00711)	(0.00825)	(0.0114)	(0.0141)
Worker Fixed Effects			No	Yes	No	Yes
Firm Fixed Effects			No	No	Yes	Yes
R^2			0.0475	0.0948	0.397	0.577
Number of Workers			12924	10063	11180	7582
Number of Firms			43466	42948	13153	11922
Number of Observations			72187	70834	42682	37936

Note: The omitted review type is Bad Review. Firms are split by the median number of jobs. 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

2.6 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 2.4 reproduces Figure 2.2. Figure 2.5 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. One hypothesis is that lower productivity workers choose to leave the labor market rather than accept low wages.[18] 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 2.6 shows a hypothetical example that is consistent with the 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 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

[18] It is also possible that higher productivity workers choose to leave instead.

Figure 2.4: Conditional Variance

Figure 2.5: Unconditional Variance

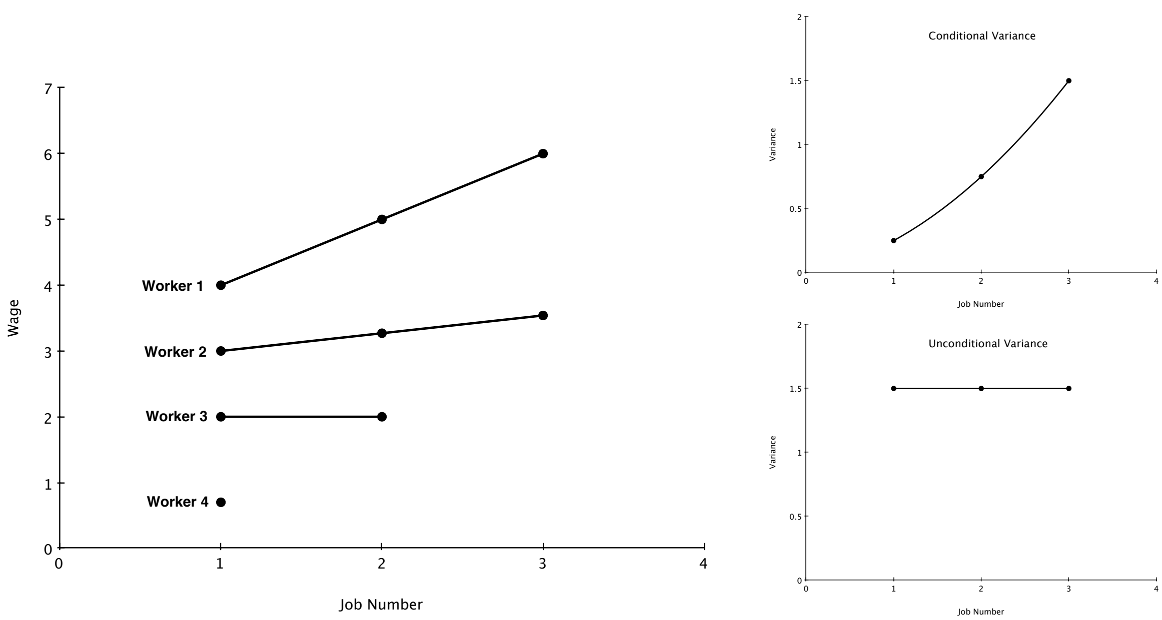
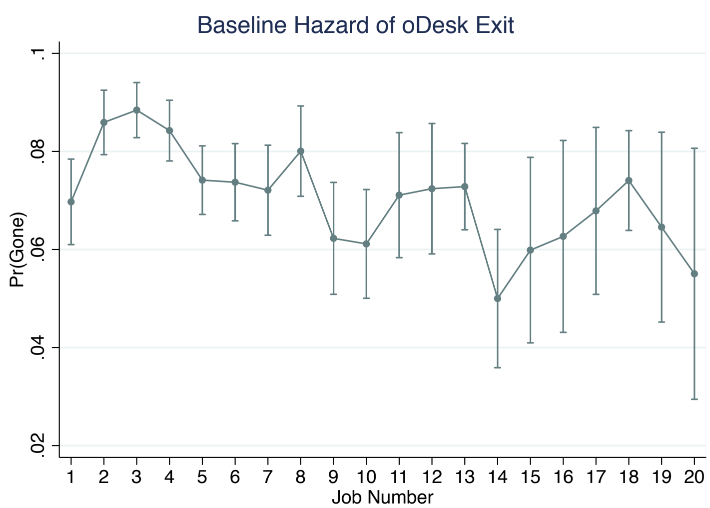


Figure 2.6: Wage Variance Example

characteristics. In particular, I allow the overall hazard rate to depend on all the variables from Equation 2.7, 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:

$$\lambda_{it} = \lambda_0(t) * exp\left(\beta_t ReviewType_{it} + \gamma_t AvgScore_{it} + \boldsymbol{\delta X_{it}} + \boldsymbol{\mu J_{it}}\right) \tag{2.8}$$

where both β_t and γ_t vary with t. Figure 2.7 shows the baseline hazard.[19] The probability of leaving the online labor market gradually declines over time.[20] This means that workers are less likely to leave the longer they stay in the market.



Note: 95% confidence bands shown.

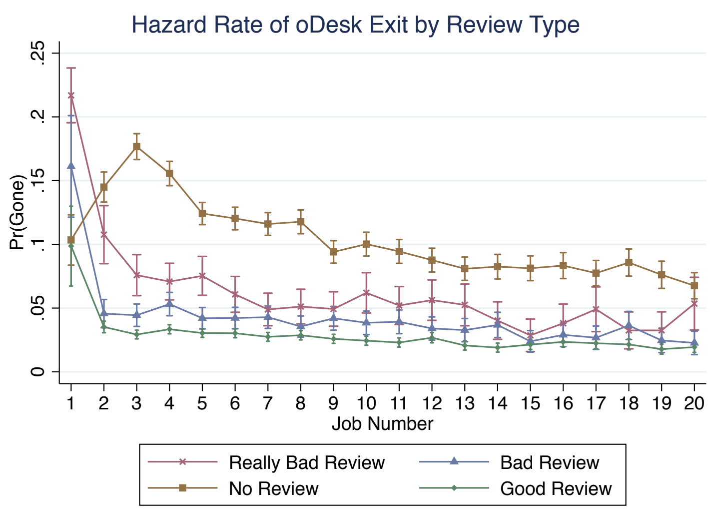
Figure 2.7: Baseline Hazard

To investigate the effect of reviews on attrition, Figure 2.8 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

[19]I 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.

[20]$p < 0.01$ for the slope of the estimated coefficients.

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 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.

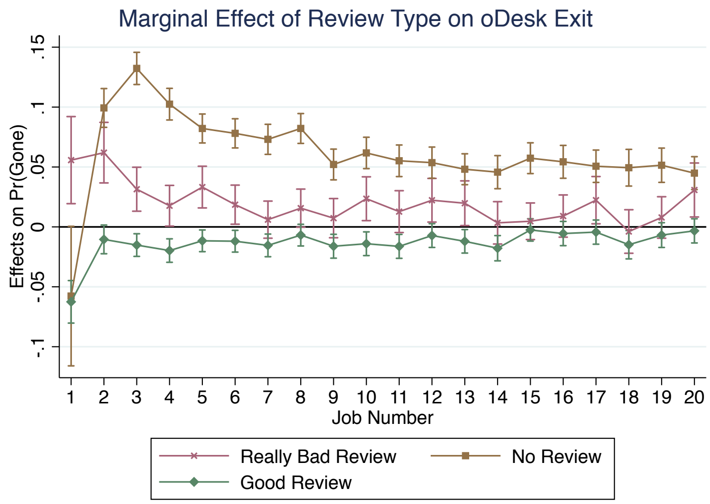


Note: 95% confidence bands shown.

Figure 2.8: Hazard Rate by Review Type

Figure 2.9 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.

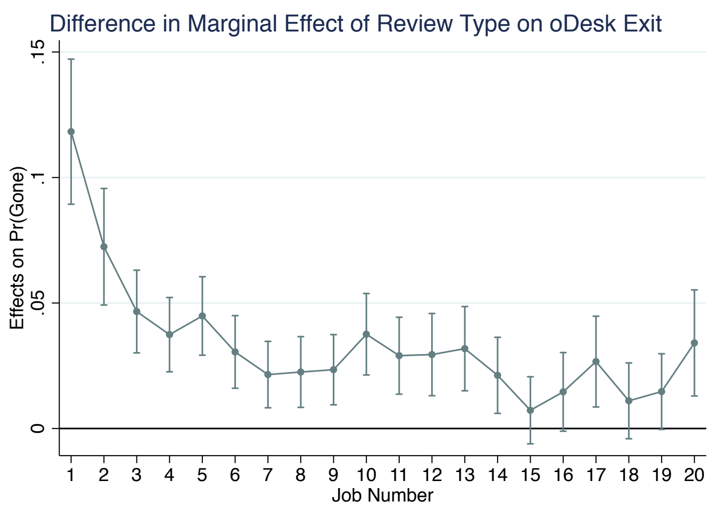
To test this convergence, Figure 2.10 plots the difference between the marginal effects of a good review and a really bad review. While the difference does not go to zero in the



Note: 95% confidence bands shown.

Figure 2.9: Marginal Effects of Review Type

first 20 jobs, there is a clear decline in the differences between types of review.[21]



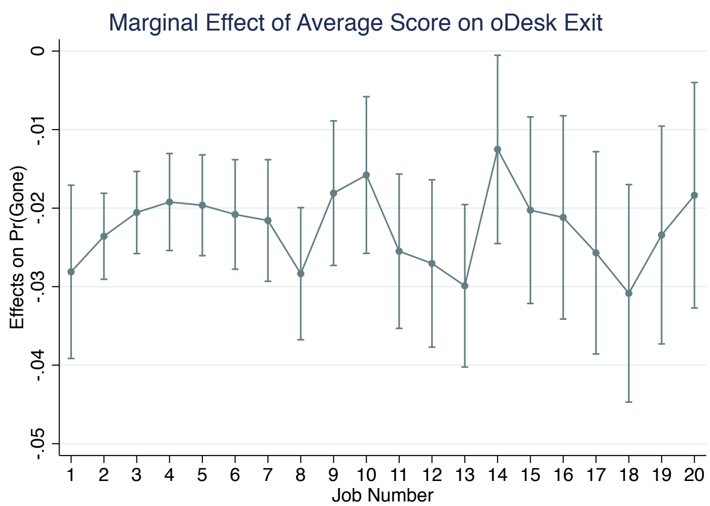
Note: 95% confidence bands shown.

Figure 2.10: Difference in Marginal Effects of Review Type

Finally, Figure 2.11 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

[21] $p < 0.01$ for the slope of the estimated coefficients.

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.



Note: 95% confidence bands shown.

Figure 2.11: Marginal Effect of Score

The above results provide additional evidence that performance reviews are providing information to the market.[22] 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. This also means that the wage results from the previous section are attenuated, since some workers who receive poor reviews exit the market rather than receive low wages. 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 investigate a possible mechanism that would induce firms to leave informative reviews.

[22]Appendix A.9 provides additional evidence that reviews affect the time to find a new job.

2.7 Why Do Employers Leave Informative Reviews?

As pointed out in the Introduction, there are a number of different reasons we might expect biased reviews in these new markets. In online labor markets in particular, there is a high percentage of repeat transactions. 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.[23] 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

[23]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.

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.

2.7.1 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.[24] 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.[25] 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.[26]

2.7.2 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 2.12 shows the decision tree.

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

[25] 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.

[26] 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.

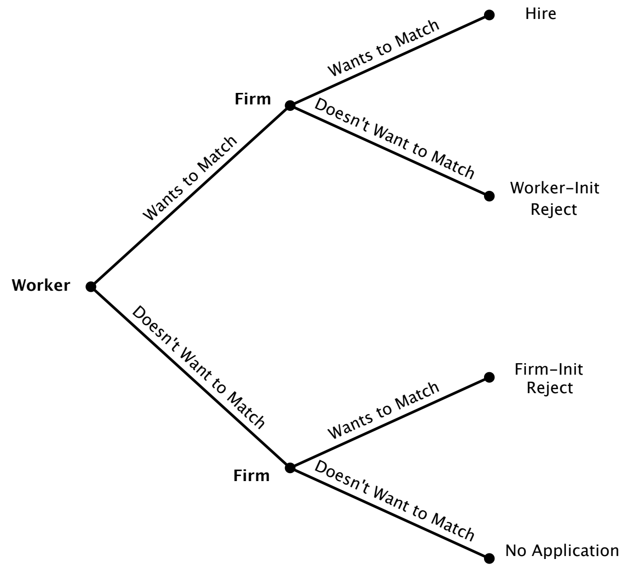


Figure 2.12: Decision Tree for Contract Renewals

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:

$$WantToMatch = \mathbb{1}\{x'\beta + \epsilon_w > 0\} \tag{2.9}$$

while the firm's decision is:

$$WantToMatch = \mathbb{1}\{z'\gamma + \epsilon_f > 0\} \tag{2.10}$$

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 2.7 for both the workers' and firms' decisions. The likelihood function for the full model is given in Equation 2.11.

$$f(outcome|\beta,\gamma) = \begin{cases} \Phi(x'\beta)\Phi(z'\gamma) & \text{if Hired} \\ \Phi(x'\beta)(1-\Phi(z'\gamma)) & \text{if Worker App, Reject} \\ (1-\Phi(x'\beta))\Phi(z'\gamma) & \text{if Client App, Reject} \\ (1-\Phi(x'\beta))(1-\Phi(z'\gamma)) & \text{if No App} \end{cases} \tag{2.11}$$

where Φ is the standard normal CDF.

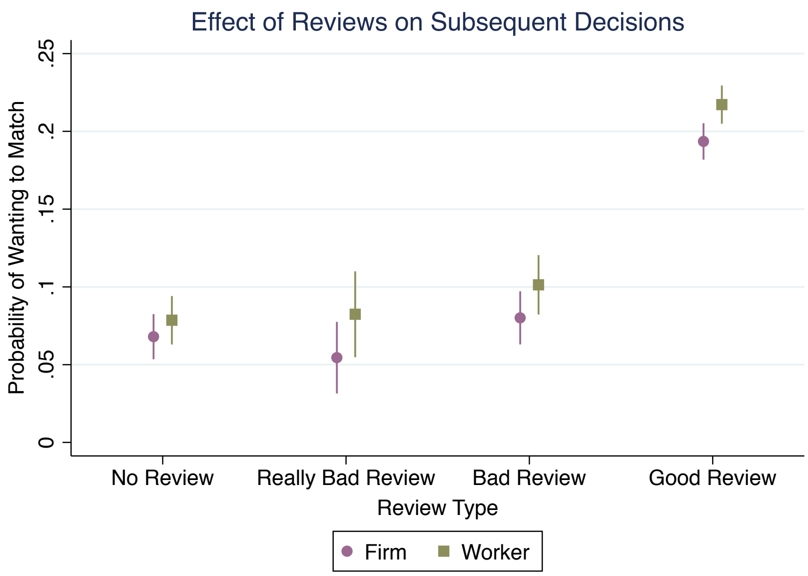
2.7.3 Results

Figure 2.13 shows predicted contract renewal rates based on maximum likelihood estimates of Equation 2.11.[27] 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

[27] Table A.3 in Appendix A.9.1 provides the coefficients.

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 are unlikely to be biased reviews.



Note: 95% confidence bands shown.

Figure 2.13: Effect of Reviews on Subsequent Decisions

2.8 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)?

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 2.14 shows the overall density of the scores and Figure 2.15 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.

2.8.1 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 2.7 and interact the review types with the comment types. Recall from Table 2.2 that relative to a good

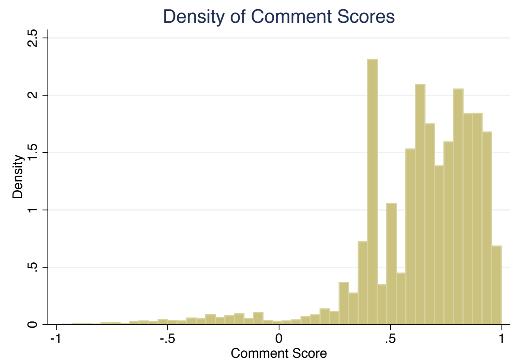
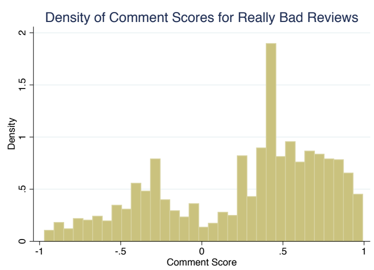
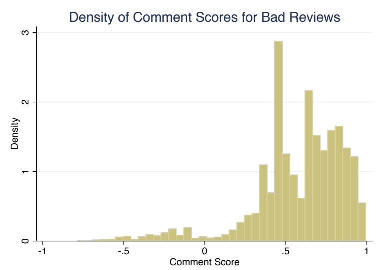


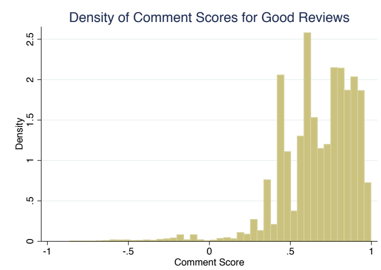
Figure 2.14: Density of Review Comment Scores



Really Bad Review Score



Bad Review Score



Good Review Score

Figure 2.15: Density of Comment Scores by Review Type

review, a bad review significantly decreased the next wage. The results in Table 2.5 are relative to a good review with any type of comment.[28] 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.

[28]I 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 2.5: Effect of Review Comments on Wage Changes

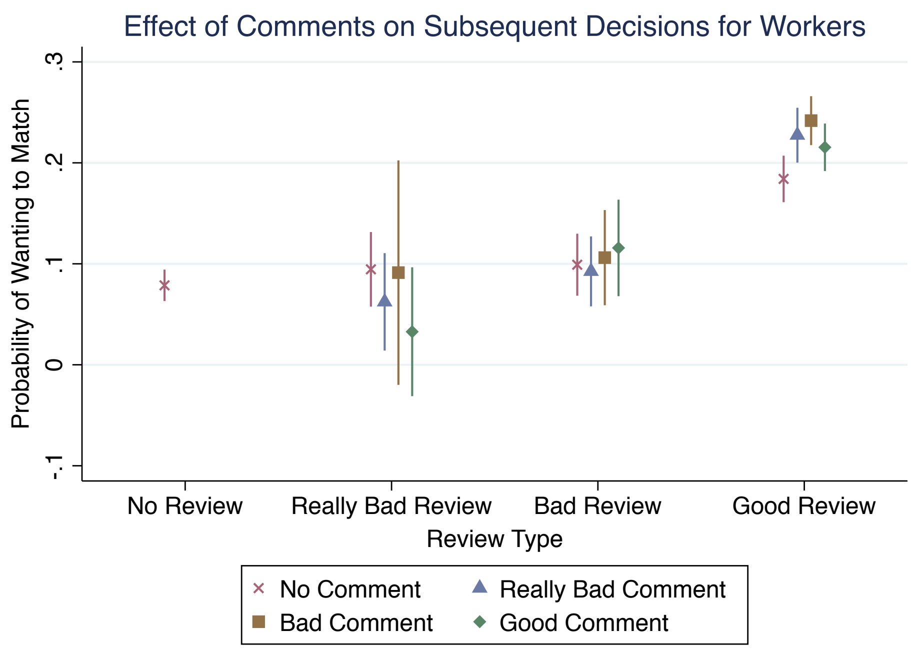
		(1)	(2)	(3)	(4)
Really Bad Review	Really Bad Comment	-0.137^{+}	-0.0852**	-0.0813*	-0.111*
		(0.0787)	(0.0295)	(0.0402)	(0.0492)
	Bad Comment	-0.00938	-0.0132	-0.0296	-0.0198
		(0.0902)	(0.0444)	(0.0591)	(0.0581)
	Good Comment	-0.202^{+}	-0.0730	-0.0965	-0.0974
		(0.116)	(0.0612)	(0.0670)	(0.0765)
	No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

2.8.2 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 2.16 shows the results for workers. There is no significant difference in workers' response to different comment types. This 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 2.16: Effect of Comments on Subsequent Decisions for Workers

2.9 Conclusion

Worker's scores on oDesk are voluntary, public, subjective, performance reviews. To date, these types of reviews have received very little theoretical or empirical attention in labor markets, in part because they were not common before the advent of online labor markets. With the rapid expansion of platforms like oDesk, however, understanding whether and how these reviews work is of increasing importance. Despite incentives for firms to bias such 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 a possible mechanism 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. This mechanism results 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.

2.10 Permissions and Attributions

Chapter 2 benefited greatly from feedback from Peter Kuhn.[29]

[29]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).

Chapter 3

How Competitive are Online Labor Markets? Evidence from oDesk

3.1 Introduction

In influential work, Card et al. (2016) (CCHK) have argued that labor market frictions (modeled as monopsony power) play important roles in explaining the wage distribution in national labor markets. Other modeling approaches based on search and matching, (e.g. Mortensen and Pissarides (1999)), come to similar conclusions about the role of labor market frictions in generating wage distributions. Recently, a number of innovations in labor markets, starting with job boards and now more recently with online labor markets, have tried to reduce these frictions by making it easier for workers and firms to match, and making more information available to both parties. Have these new types of labor markets succeeded in reducing market frictions? I study this question by comparing an online labor market, oDesk, to offline labor markets. First, I use data from oDesk to estimate a two-way fixed effects model in the style of Abowd, Kramarz and Margolis (1999) (AKM). These models allow for a decomposition of worker and employer-specific wage premiums. I compare my estimates to the existing literature on AKM estimates for the United States and find that online labor markets have a higher variance of employer-specific wage premiums. I then use the imperfect competition model from CCHK to show that the underlying parameters governing market frictions can be mapped to these employer-specific wage premiums. This suggests that online labor markets are more competitive than traditional labor markets.

There is a significant body of literature in labor economics that tries to explain the variance of worker wages. Much of the search literature has recognized the importance of frictions in search and matching between workers and firms in generating a distribution of wages, even for equivalent workers (Mortensen and Pissarides, 1999; Rogerson, Shimer and Wright, 2005; Pissarides, 2011). A second strand of literature tries to decompose the variance of wages into worker and firm components. In the pioneering work of AKM, they use matched employer-employee data from France to separately identify the contributions of workers and firms to wage variation. This style of two-way fixed effects model has been used to study the labor markets of a number of different countries. All of these studies use linked employer-employee administrative data which includes information on wages and some worker characteristics. Table 3.1 provides a summary of recent results.[1] Between 15-30% of the variation in wages is explained by variation in the firm effects. There is substantial heterogeneity across the countries, with the United States having by far the largest variance of firm effects. As we will see from the model, this suggests the US labor market is the most competitive of the countries listed. For a given distribution of firm productivity, a higher variance of firm effects implies that workers face fewer frictions in moving between employers in response to those productivity differentials and that wages are closer to firms' marginal product.

[1] All of these studies look across a number of different industries, however Abowd et al. (2014) estimate an AKM model within industries and the results are similar.

CCHK provide a micro-founded model of imperfect competition that generates wages that are consistent with the assumptions of AKM. I use the CCHK model to map the variance of firm productivity, the degree of labor market frictions, and the workers' outside options to the theoretical firm fixed effects. I then use estimates of the firm fixed effects using data from oDesk, as well as existing US estimates from Abowd, Lengermann and McKinney (2002) to compare three different labor markets: the oDesk market for US workers, the oDesk market for developing country workers, and the offline US market. These comparisons, combined with the assumptions in the model, let me infer differences in competitiveness across markets. I find evidence that the online labor market is more competitive than the offline US market. I also find evidence that the outside options of workers on oDesk are lower than in traditional labor markets, which potentially explains the growth of these online labor markets.

The rest of the paper is organized as follows. Section 2 discusses the empirical framework. Section 3 describes the data and Section 4 presents the empirical results. Section 5 lays out the imperfect competition model of CCHK and Section 6 maps the empirical results to implications for competitiveness.

3.2 Empirical Framework

The empirical analysis will follow AKM by specifying a model of log wages with additive worker and firm effects. Specifically, in a given market, I will model the log wage of person i, working for firm j, at job number t as:

$$\ln w_{ijt} = \alpha_i + \psi_j + X_{it}'\theta + \epsilon_{ijt} \tag{3.1}$$

where X_{it} is a vector of time varying controls (e.g. worker experience and year dummies), α_i is the worker fixed effect, ψ_j is the firm fixed effect and ϵ_{ijt} is an unobserved error term. Estimates from this model can then be used to decompose wages according to the following formula:

$$\begin{aligned} Var(\ln w_{ijt}) &= Var(\alpha_i) + Var(\psi_j) + Var(X'_{it}\theta) + Var(\epsilon_{ijt}) \\ &+2Cov(\alpha_i, \psi_j) + 2Cov(\alpha_i, X'_{it}\theta) + 2Cov(\psi_j, X'_{it}\theta) \end{aligned} \tag{3.2}$$

Estimating this model for each labor market allows us to compare the relative variation in firm fixed effects.

3.3 Data

3.3.1 Online Labor Markets

Online labor markets are a growing proportion of the world economy.[2] 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, 2015). 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. 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

[2] Most of this description comes from Wood-Doughty (2017)

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 then 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.[3] oDesk facilitates matches by allowing workers and firms to search for each other with very detailed filters.

3.3.2 Data Description

The data consists of a sample of oDesk workers who were active in the administrative job category between January 1, 2015 and April 25, 2015. I observe every job these 11,676 workers have done on oDesk from when they first joined, (some as early as 2007), through April 25, 2015, with the review they receive. I choose to look at administrative jobs because they are more homogenous than other categories and the majority of them are hourly jobs.[4] I also observe each worker's 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 3.2 provides summary statistics for the workers. I split my sample into workers from developing countries and those from the United States. Half of all workers from developing countries 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, and personal assistant jobs. These jobs are generally low-skilled and pay a relatively low hourly wage.[5] There are significant wage differences between the workers

[3]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.

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

[5]However, the hourly wage is comparable to the average hourly wage in India and the Philippines.

from developing countries and those from the United States. The average (partial) career is ten jobs with each job being a fairly significant time commitment: almost a month long and 40 hours per week. The average review score is around a 4.7 out of 5. While this type of market is different from the standard labor market in developed countries, it is not all that different from temporary staffing agencies, which account for around 2% of the U.S. labor market (American Staffing Association, 2017; Bureau of Labor Statistics, 2017).

For each job a worker completes, I also observe information about the hiring firm. Table 3.3 provides summary statistics for these firms again split by which market they are in. The first column provides a summary of all the firms in the sample. For around half of them, I have additional data on firm characteristics, which are displayed in the second column. Almost all of the firms are located in OECD countries, with the majority in the United States. The average review a firm receives after a completed job is around a 4.7 out of 5. The average number of jobs a firm has hired is quite high, although this is skewed by a few very frequent hirers (the median number of jobs is 29). For a subset of firms, I also have data on the average review they left for workers and the average amount they pay per job.

As mentioned above, in order to identify the fixed effects, the network of workers and firms must be connected. Furthermore, the precision of the fixed effects estimates depends on the number of observations for each worker and firm. The more moves each worker makes and the more workers each firm hires, the more precise the estimates. Table 3.4 provides summary statistics for worker and firm moves in the sample. Contrary to traditional labor markets, most firms in online labor markets are quite small, in that they only employ a few workers in this market. This is consistent with traditional firms going to an online setting to hire specialized workers for a specific task.

3.4 Empirical Results

To decompose the variance of wages in an online labor market, I estimate Equation 3.1 and then decompose the variance according to Equation 3.2. Table 3.5 presents the results. The variance of wages is slightly higher for workers from developing countries than for workers from the US. The variance of worker effects are very similar, but the variance of firm effects for US workers is much higher. Comparing these results to the US results in Table 3.1 suggests that the variance of firm effects for developing countries is slightly higher than the traditional US labor market, but the variance of firm effects for oDesk US workers is significantly higher. It is important to keep in mind that the oDesk market is only for administrative jobs, and to the extent that cross-industry heterogeneity is an important component of overall firm heterogeneity, the estimates of variance in the oDesk markets will be biased downwards relative to the traditional labor markets. This suggests there are actually larger differences in the variances than we observe.

3.4.1 Controlling for Job Heterogeneity

One explanation for the large variance of employer-specific wage premiums in online labor markets is that different firms are hiring for different types of jobs, and the wage premium just reflects heterogeneity in job types. Torres et al. (2013) use administrative data from Portugal and find that job titles explain almost 10% of the wage variation. While Portugal has a very standardized way of naming jobs, jobs titles in online labor markets are determined by the hiring firm and are thus free-form text.[6] I attempt to control for job heterogeneity in five different ways. First, each administrative job in this online labor market is further classified according to a subcategory. These categories

[6]The titles in my dataset are much noisier than in standard administrative datasets. Marinescu and Wolthoff (2016) use job titles from CareerBuilder.com, which has more standard job titles, and find that the titles explain almost 90% of the variation in wages. In my sample, all of the techniques combined only explain 35% of the variation in wages.

include "Admin Support", "Sales & Marketing", and "Web, Mobile & Software Development." I include indicator variables for each of these subcategories. Second, there is some job-specific information available including the target skill level, target duration of the job, and the number of applicants. For the last three classifications, I use the text from each job title. First, I use latent Dirichlet allocation (LDA) to classify each title. LDA is a Bayesian model that generates topics based on similar words (Blei, Ng and Jordan, 2003). It then classifies each title as one of the topics. Next, I include a number of indicators for important words in the titles. To do this, I construct a separate dataset that includes job titles and wages. I then run a cross-validated LASSO of the job wage on the different words in the title (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. Finally, I create a neural network in the style of Mikolov et al. (2013). This creates a 500-element vector for each title where similar titles have similar values. I then include these vectors in the regression to control for differences between jobs. Table 3.6 presents the results. Columns 1 and 6 are the baseline specifications for developing countries and the US markets, respectively. Columns 2 through 5 sequentially add additional controls for job heterogeneity for the developing country market. There does not seem to be much change in the variance of firm fixed effects, especially when we add in the most detailed controls. Due to the smaller sample size of the US-based oDesk market, I can only do the first two controls. Here we do see a significant decrease in the variance of firm fixed effects. However, these variances are still larger than both the other oDesk market and the traditional US market, so the overall results still hold.

3.5 Theoretical Framework

To formalize the relationship between wages, workers, and firms, I follow the imperfect competition model of Card et al. (2016). Worker i in labor market m has indirect utility from working for firm j:

$$u_{imj} = \beta ln(w_{mj} - b_{im}) + a_{mj} + \epsilon_{imj} \tag{3.3}$$

where w_{mj} is the wage paid by firm j in market m, b_{im} is the worker's outside option, a_{mj} is observable firm amenities, and ϵ_{imj} is an unobservable match preference.[7] Assuming ϵ_{imj} is distributed type I extreme value and there are a large number of firms, leads to the following firm-specific labor supply functions:

$$lnN_{mj}(w_{mj}) = ln(\mathcal{N}\lambda) + \beta_m ln(w_{mj} - b_m) + a_{mj} \tag{3.4}$$

where $\mathcal{N}$ is the number of workers and λ is constant across all firms. The parameter of interest is β_m which governs the relative importance to workers of wages compared to firm amenities. As $\beta_m \to \infty$ the supply functions become perfectly elastic and we approach a competitive labor market.

Assume firms have a linear production function and face constant price, P_{mj}:

$$Revenue_{mj} = P_{mj}T_{mj}\sum_i \theta_{mi} \tag{3.5}$$

where T_{mj} is firm productivity and θ_{mi} is an individual worker's productivity, then solving

[7]Firm amenities in an online labor market may include availability, communication, and clarity of the task.

for the profit maximizing wages gives:

$$w_{imj} = \frac{b_{im}}{1+\beta_m} + \frac{\beta_m}{1+\beta_m}\nu_{mj}\theta_{mj} \tag{3.6}$$

where $\nu_{mj} \equiv T_{mj}P_{mj}$ is marginal labor productivity. Assume that a worker's outside option is a function of their productivity: $b_{im} = \theta_{im}b_m$, then:

$$w_{imj} = \frac{\theta_{im}b_m}{1+\beta_m} + \frac{\beta_m}{1+\beta_m}\nu_{mj}\theta_{mj} = \frac{\theta_{im}b_m}{1+\beta_m}(1+\frac{\beta_m\nu_{mj}}{b_m}) \tag{3.7}$$

Taking logs gives us:

$$\ln w_{imj} = \ln(\frac{\theta_{im}b_m}{1+\beta_m}) + \ln(1+\frac{\beta_m\nu_{mj}}{b_m}) = \alpha_{im} + \psi_{mj} \tag{3.8}$$

which is additively separable in worker and firm effects within a given labor market and can be estimated via AKM.

3.6 Implications for Competitiveness

The firm fixed effects relate our three parameters of interest: β_m, b_m, and $\bar{\nu}_{mj}$, where $\nu_{mj} \in [b_m, \bar{\nu}_m]$[8]. We cannot separately identify the level of these parameters using only wage data from these markets. However, we can relate these underlying parameters to observable differences across markets. Combined with some plausible assumptions about differences between markets, we can rank order the parameters of interest across the three markets. The key observable difference between the markets is the variance of the

[8] Since firms will not operate if their marginal productivity is less than workers' outside options, the lower bound of productivity in the market is b_m, while we are allowing the upper bound, $\bar{\nu}_m$ to vary in different markets.

firm fixed effects:[9]

$$Var(\psi_{mj}) = Var(\ln(1 + \frac{\beta_m \nu_{mj}}{b_m})) \tag{3.9}$$

While we do not have a nice closed form solution to $Var(\psi_j)$, we can sign the derivatives. Taking the derivative with respect to the competitiveness of the market:

$$\frac{\partial Var(\psi_{mj})}{\partial \beta_m} > 0 \tag{3.10}$$

The more competitive the market, the more variation in firm fixed effects. In a perfectly competitive market, firms pay their marginal product, so the distribution of firm fixed effects is the same as the distribution of firm productivity. As firms have more market power, wages are compressed, so the variance of firm effects declines.

Taking the derivative with respect to the upper bound of firm productivity:

$$\frac{\partial Var(\psi_{mj})}{\partial \bar{\nu}_m} = \frac{\partial Var(\psi_{mj})}{\partial Var(\nu_{mj})} * \frac{\partial Var(\nu_{mj})}{\partial \bar{\nu}_m} > 0 \tag{3.11}$$

As the most productive firm in the market gets more productive, the overall distribution gets more spread out, holding the bottom of the distribution fixed. Thus the variance of firm fixed effects increases.

Taking the derivative with respect to the outside option:

$$\frac{\partial Var(\psi_{mj})}{\partial b_m} = \frac{\partial Var(\psi_{mj})}{\partial b_m} + \frac{\partial Var(\psi_{mj})}{\partial Var(\nu_{mj})} * \frac{\partial Var(\nu_{mj})}{\partial b_m} < 0 \tag{3.12}$$

which is an abuse of notation. There are two effects here, both of which are negative. First, there is a direct effect, by which increasing the outside option reduces the relative pay gap between the market wage and the outside option. Second there is an indirect

[9] And an important value in much of the AKM literature.

effect through $Var(\nu_{mj})$, since the outside option is the lower bound of firm productivity in this market. Increasing this lower bound, holding the rest of the distribution constant, will decrease the variance of firm effects.

Taking these effects together, if we compare the variance of firm effects between two markets, we can determine the relationship between these underlying parameters. For example, consider two markets: A and B. If we find that the variance of firm effects in market A is greater than the variance of firm effects in market B we can conclude that:

$$Var(\psi^A) > Var(\psi^B) \Rightarrow \beta^A > \beta^B \text{ and/or } b^A < b^B \text{ and/or } \bar{\nu}^A > \bar{\nu}^B \tag{3.13}$$

If we make some reasonable assumptions about the relationships between some of the parameters, we can potentially narrow down to learning about one of the parameters. Given that we have some intuition about the different markets of study in this paper, we can make some reasonable assumptions.

Comparing the US labor market with the online labor markets for the US (USO) and developing country workers (Dev), we can assume that:

$$b^{Dev} < b^{USO} = b^{US} \tag{3.14}$$

The outside options in developing countries are worse than in the US. Both US markets have the same outside options (since this is net of worker ability).

$$\bar{\nu}_j^{Dev} = \bar{\nu}_j^{USO} < \bar{\nu}_j^{US} \tag{3.15}$$

The firms are the same in oDesk since they are operating in the same online market.

There are better firms in the US offline market.

$$\beta^{US} \lesseqgtr \beta^{Dev} \lesseqgtr \beta^{USO} \tag{3.16}$$

We are ambivalent about the relative competitiveness between these markets. The goal of this paper is to sign these relationships.

Then combining these assumptions with the comparisons of variances, gives us the following predictions: Comparing US oDesk to US, if

$$Var(\psi_j^{USO}) > Var(\psi_j^{US}) \Rightarrow \beta^{USO} > \beta^{US} \tag{3.17}$$

or if

$$Var(\psi_j^{USO}) < Var(\psi_j^{US}) \Rightarrow \bar{\nu}_j^{USO} < \bar{\nu}_j^{US} \tag{3.18}$$

A higher variance in the online market suggests that competitiveness is driving the difference. A higher variance in the offline market suggests that differences in firm productivity are more important.

Comparing the two oDesk markets, if

$$Var(\psi_j^{Dev}) > Var(\psi_j^{USO}) \Rightarrow b^{Dev} < b^{USO} \tag{3.19}$$

or if

$$Var(\psi_j^{Dev}) < Var(\psi_j^{USO}) \Rightarrow \beta^{Dev} < \beta^{USO} \tag{3.20}$$

A higher variance in the developing country market suggests that a lower outside option for those workers is driving the differences. A higher variance in the US market means that there are competitiveness differences.

Comparing developing countries to US, if

$$Var(\psi_j^{Dev}) > Var(\psi_j^{US}) \Rightarrow b^{Dev} < b^{US} \text{ and/or } \beta^{Dev} > \beta^{US} \tag{3.21}$$

or if

$$Var(\psi_j^{Dev}) < Var(\psi_j^{US}) \Rightarrow \bar{\nu}_j^{Dev} < \bar{\nu}_j^{US} \tag{3.22}$$

A higher variance in the developing country market means that either there are outside option differences or competitiveness differences, or, most likely, both together. A higher variance in the offline US market means that the high end of firm productivity is driving the differences.

Combining the empirical results in Section 3.4 with the above assumptions allows me to infer difference in competitiveness across these markets. From Tables 3.1 and 3.5 we observe:

$$Var(\psi^{US}) < Var(\psi^{Dev}) < Var(\psi^{USO}) \tag{3.23}$$

Turning to the theoretical predictions, comparing US oDesk to US:

$$Var(\psi_j^{US}) < Var(\psi_j^{USO}) \Rightarrow \beta^{US} < \beta^{USO} \tag{3.24}$$

which suggests the US oDesk market is more competitive than the offline market. Comparing the two oDesk markets:

$$Var(\psi_j^{Dev}) < Var(\psi_j^{USO}) \Rightarrow \beta^{Dev} < \beta^{USO} \tag{3.25}$$

which suggests there is more evidence for a difference in competitiveness than a difference in outside options. And finally, comparing developing countries to US, if

$$Var(\psi_j^{Dev}) > Var(\psi_j^{US}) \Rightarrow b^{Dev} < b^{US} \text{ and/or } \beta^{Dev} > \beta^{US} \tag{3.26}$$

which suggests that the oDesk market is more competitive but that the workers have lower outside options.

Taken together, the evidence suggests that this online labor market is more competitive than the traditional labor market in the US. There is also evidence that workers in developing countries have lower outside options, and therefore are willing to work for lower wages. This is one possible reason why firms are willing to move to these more competitive marketplaces. Finally, it is important to note from Table 3.1 that the US had the largest variance of firm effects, so these results will also hold in comparisons with the other countries listed.

3.7 Conclusion

Online labor markets are an important part of the growing gig economy. These new forms of labor markets have promised a more competitive marketplace with fewer frictions. However, there is not yet much empirical evidence to support this claim. This paper uses data from an online labor market to compare these markets to a traditional labor market. I decompose the wage variance using an AKM-style model and find that online labor markets have a higher variance of firm fixed effects. I then use a theoretical model of imperfect competition to map the variance of firm fixed effects to the degree of frictions in a given labor market. My results imply that online labor markets are more competitive than traditional markets. I also find evidence that workers' outside options are worse in these online markets, which suggests a reason for why firms may want to move to these new alternative work arrangements.

3.8 Permissions and Attributions

Chapter 3 benefited greatly from feedback from Peter Kuhn.

Table 3.1: Summary of Two-Way Fixed Effects Results

Country	Paper	Variance of Log Wages	Variance of Worker Effects	Variance of Firm Effects
France	Abowd, Creecy and Kramarz (2002)	0.270	0.207	0.081
US	Abowd, Lengermann and McKinney (2002)	0.776	0.697	0.131
New Zealand	Maré and Hyslop (2006)	0.116	0.058	0.010
Austria	Gruetter and Lalive (2009)	0.224	0.148	0.083
West Germany	Card, Heining and Kline (2012)	0.249	0.127	0.053
Portugal	Torres et al. (2013)	0.309	0.074	0.056
Italy	Macis and Schivardi (2016)	0.116	0.058	0.017

Note: The results in this table are not directly comparable, given sample selection and choice of time-varying variables

Table 3.2: Summary Statistics for Workers

	Developing Countries	US
Master's Degree	0.0797	0.0494
	(0.271)	(0.217)
Bachelor's Degree	0.518	0.385
	(0.500)	(0.487)
Associate's Degree	0.0114	0.0392
	(0.106)	(0.194)
Other Degree	0.391	0.526
	(0.488)	(0.500)
Number of Outside Experiences	2.636	3.578
	(2.510)	(3.075)
Number of oDesk Exams	1.947	1.986
	(1.250)	(1.323)
Average Wage ($)	4.057	11.04
	(2.981)	(5.448)
Average Job Duration (Hours)	181.0	121.1
	(306.2)	(210.2)
Average Job Duration (Days)	26.57	30.48
	(47.73)	(48.83)
Number of Jobs	11.12	9.775
	(10.65)	(8.834)
Average Review Score	4.650	4.777
	(0.457)	(0.358)
N	10503	1173

Note: Means reported, standard deviations in parentheses.

Table 3.3: Summary Statistics for Firms

	Developing Countries		US	
	Full Sample	Extra Covariates	Full Sample	Extra Covariates
High Income Country: OECD	0.926	0.926	0.948	0.946
	(0.261)	(0.261)	(0.223)	(0.226)
High Income Country: Non-OECD	0.0336	0.0336	0.0283	0.0296
	(0.180)	(0.180)	(0.166)	(0.170)
Upper Middle Income Country	0.0142	0.0132	0.0114	0.00878
	(0.118)	(0.114)	(0.106)	(0.0933)
Lower Middle Income Country	0.0221	0.0223	0.0109	0.0132
	(0.147)	(0.148)	(0.104)	(0.114)
Low Income Country	0.00325	0.00397	0.000544	0
	(0.0569)	(0.0629)	(0.0233)	(0)
Average Review Score Received	4.737	4.771	4.750	4.815
	(0.632)	(0.414)	(0.688)	(0.326)
Number of Jobs	61.88	65.92	130.9	117.1
	(507.2)	(123.8)	(1579.2)	(283.0)
Average Review Score Left		4.510		4.593
		(0.587)		(0.517)
Average Total Payment per Job ($)		1303.3		3280.7
		(4385.8)		(8001.1)
N	19706	9561	1838	911

Note: Means reported, standard deviations in parentheses.

Table 3.4: Movement by Workers and Firms

	Developing Countries		US	
	Worker Moves	Firm Moves	Worker Moves	Firm Moves
Mean	5.50	2.93	2.74	1.75
Minimum	1	1	1	1
25th Percentile	2	2	1	1
50th Percentile	4	2	2	1
75th Percentile	7	3	3	2
Maximum	61	179	35	47

Table 3.5: Decomposition of Wage Variance

	Developing Countries	US
Variance of Log Wages	0.54	0.48
Variance of $X'\beta$	0.036	0.026
Variance of Worker Effects (α_i)	0.25	0.24
Variance of Firm Effects (ψ_j)	0.15	0.26
Variance of Residuals	0.056	0.029
2*Cov($X'\beta$, α_i)	-0.0095	0.00022
2*Cov($X'\beta$, ψ_j)	0.020	0.0068
2*Cov(α_i, ψ_j)	0.036	-0.078
Number of Workers	10503	1173
Number of Firms	19706	1838
Number of Observations	77288	5399

Note: AKM decomposition of the variance of log wages. Regression includes indicators for years and a quadratic in worker experience.

Table 3.6: Decomposition of Wage Variance Controlling for Job Titles

	Developing Countries					US		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variance of Log Wages	0.536	0.536	0.536	0.554	0.554	0.480	0.480	0.480
Variance of $X'\beta$	0.036	0.045	0.048	0.051	0.055	0.026	0.029	0.039
Variance of Worker Effects (α_i)	0.252	0.232	0.226	0.222	0.218	0.240	0.316	0.317
Variance of Firm Effects (ψ_j)	0.147	0.127	0.126	0.149	0.148	0.256	0.162	0.174
Variance of Residuals	0.056	0.054	0.053	0.049	0.048	0.029	0.029	0.027
Job Info		X	X	X	X		X	X
Job Categories		X	X	X	X		X	X
LDA Topics			X	X	X			X
Job Title Dummies				X	X			
Word2Vec Vectors					X			
Number of Workers	10503					1173		
Number of Firms	19706					1838		
Number of Observations	77288					5399		

Note: AKM decomposition of the variance of log wages. Regression includes indicators for years and a quadratic in worker experience.

Appendix A

Appendix to Chapter 2

A.1 Symmetric Employer Learning Specification

Table A.1 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%.

Table A.1: 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:[1]

$$r_{it} = \arg\max_{\tilde{r}_{it}} \Bigg(\mathbb{E}[\theta_i|s_i, r_{i1}, ..., y_{it}] - \mathbb{E}[\theta_i|s_i, r_{i1}, ..., \tilde{r}_{it}] \Bigg) f(\tilde{r}_{it}|c^H) - h(\tilde{r}_{it}|c^R) - l(y_{it}, \tilde{r}_{it}|c^L) \tag{A.1}$$

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:

$$w_{it+1} = \mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it}] + \eta_{ijt+1} \tag{A.2}$$

$$\begin{aligned} r_{it} = \arg\max_{\tilde{r}_{it}} \Bigg(\mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it-1}, y_{it}] - \mathbb{E}[\theta_i|x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}] \Bigg) &\Phi(\alpha + c^H r_{it}) \\ &+ c^R r_{it} - c^L(y_{it} - r_{it})^2 \end{aligned} \tag{A.3}$$

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

Consider four possible examples:

No Costs

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

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = -\frac{\partial\, \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) < 0 \Rightarrow r_{it} = 1 \tag{A.4}$$

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.

Infinite Rehire Costs

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

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = -\frac{\partial\, \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha + c^H r_{it}) + \tag{A.5}$$

$$\left(\mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, y_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}] \right) c^H \phi(\alpha + c^H r_{it}) = 0 \tag{A.6}$$

$$\Rightarrow r_{it} = y_{it} \tag{A.7}$$

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.

Infinite Reputation Costs

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

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = -\frac{\partial\, \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) + c^R \Rightarrow r_{it} = 5 \tag{A.8}$$

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

Infinite Lying Costs

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

$$\frac{\partial \pi(y_{it}, \tilde{r}_{it})}{\partial \tilde{r}_{it}} = -\frac{\partial\, \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}]}{\partial \tilde{r}_{it}} \Phi(\alpha) \tag{A.9}$$

$$+2c^L \left(\mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, y_{it}] - \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it-1}, \tilde{r}_{it}] \right) = 0 \tag{A.10}$$

$$\Rightarrow r_{it} = y_{it} \tag{A.11}$$

since their 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 cross-sectional variance of wages increases with t. In an uninformative world, the variance is constant.*

Proof. (Shown for the first review to keep the algebra simple)

In an informative world,

$$w_{ij1} = \mathbb{E}[\theta_i|x_i] + \eta_{ik1}$$

$$w_{ij2} = \mathbb{E}[\theta_i|x_i, r_{i1}] + \eta_{ik2}$$

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

$$\mathbb{E}[\theta_i|x_i] = \alpha^1 + \beta^1 x_i$$

$$\mathbb{E}[\theta_i|x_i, r_{i1}] = \alpha^2 + \beta^2 x_i + \lambda_1^2 r_{i1}$$

so

$$Var(w_{ij1}) = (\beta^1)^2 Var(x_i) + Var(\eta_{ik1})$$

$$Var(w_{ij2}) = (\beta^2)^2 Var(x_i) + (\lambda_1^2)^2 Var(r_{i1}) + 2\beta^2\lambda_1^2 Cov(x_i, r_{i1}) + Var(\eta_{ik2})$$

therefore

$$Var(w_{ij1}) = (\frac{Cov(x_i, \theta_i)}{Var(x_i)})^2 Var(x_i) + \sigma_\eta^2 = Corr(x_i, \theta_i)^2 Var(\theta_i) + \sigma_\eta^2$$

and

$$Var(w_{ij2}) = \left(\frac{Var(r_{i1})Cov(x_i,\theta_i) - Cov(x_i,r_{i1})Cov(r_{i1},\theta_i)}{Var(r_{i1})Var(x_i) - Cov(x_i,r_{i1})^2}\right)^2 Var(x_i)$$
$$+\left(\frac{Var(x_i)Cov(r_{i1},\theta_i) - Cov(x_i,r_{i1})Cov(x_i,\theta_i)}{Var(r_{i1})Var(x_i) - Cov(x_i,r_{i1})^2}\right)^2 Var(r_{i1})$$
$$+2\left(\frac{Var(r_{i1})Cov(x_i,\theta_i) - Cov(x_i,r_{i1})Cov(r_{i1},\theta_i)}{Var(r_{i1})Var(x_i) - Cov(x_i,r_{i1})^2}\right)$$
$$\left(\frac{Var(x_i)Cov(r_{i1},\theta_i) - Cov(x_i,r_{i1})Cov(x_i,\theta_i)}{Var(r_{i1})Var(x_i) - Cov(x_i,r_{i1})^2}\right)Cov(x_i,r_{i1})$$
$$+\sigma_\eta^2$$

Simplifying gives

$$Var(w_{ij2}) = \frac{Var(\theta_i)}{1 - Corr(x_i,r_{i1})^2}\Big(Corr(x_i,\theta_i)^2 + Corr(r_{i1},\theta_i)^2 -$$
$$2Corr(x_i,\theta_i)Corr(r_{i1},\theta_i)Corr(x_i,r_{i1})\Big) + \sigma_\eta^2$$

so

$$Var(w_{ij2}) - Var(w_{ij1})$$
$$= \frac{Var(\theta_i)}{1 - Corr(x_i,r_{i1})^2}\Big(Corr(x_i,\theta_i)^2 + Corr(r_{i1},\theta_i)^2$$
$$-2Corr(x_i,\theta_i)Corr(r_{i1},\theta_i)Corr(x_i,r_{i1})\Big) - Corr(x_i,\theta_i)^2 Var(\theta_i)$$
$$= \frac{Var(\theta_i)}{1 - Corr(x_i,r_{i1})^2}\Big(Corr(x_i,\theta_i)^2 + Corr(r_{i1},\theta_i)^2$$
$$-2Corr(x_i,\theta_i)Corr(r_{i1},\theta_i)Corr(x_i,r_{i1}) - Corr(x_i,\theta_i)^2(1 - Corr(x_i,r_{i1})^2)\Big)$$
$$= \frac{Var(\theta_i)}{1 - Corr(x_i,r_{i1})^2}\Big(Corr(r_{i1},\theta_i)^2 - 2Corr(x_i,\theta_i)Corr(r_{i1},\theta_i)Corr(x_i,r_{i1})$$

$$+Corr(x_i,\theta_i)^2 Corr(x_i,r_{i1})^2\Big)$$
$$= \frac{Var(\theta_i)}{1-Corr(x_i,r_{i1})^2}\Big(Corr(r_{i1},\theta_i) - Corr(x_i,\theta_i)Corr(x_i,r_{i1})\Big)^2 > 0$$

In an uninformative world,

$$w_{ij1} = \mathbb{E}[\theta_i|x_i] + \eta_{ik1}$$
$$w_{ij2} = \mathbb{E}[\theta_i|x_i] + \eta_{ik2}$$

so

$$Var(w_{ij1}) = (\beta^1)^2 Var(x_i) + Var(\eta_{ik1})$$
$$Var(w_{ij2}) = (\beta^1)^2 Var(x_i) + Var(\eta_{ik2})$$

and

$$Var(w_{ij2}) - Var(w_{ij1}) = 0$$

□

A.4 Proof of Proposition 2

Proposition. *In an informative world, the marginal effect of the observable characteristic on the conditional mean of wages is strictly positive and decreasing with t. In an uninformative world, the marginal effect is strictly positive and constant.*

Proof.

In an informative world,

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

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

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

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

so

$$\frac{\partial w_{ijt+1}}{\partial x_i} = \beta$$

where β is the regression coefficient of x_i on θ_i. Since this is increasing in $Cov(\theta_i, x_i)$ and

$$Cov(\theta_i, x_i) = \gamma > 0$$

therefore

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

since β is decreasing in the number of positively correlated regressors,

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

In an uninformative world,

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

$$\frac{\partial w_{ijt+1}}{\partial \bar{r}_{it}} > 0$$

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

□

A.5 Proof of Proposition 3

Proposition. *In an informative world, the marginal effect of the worker's average review score on the conditional mean of wages is strictly positive and increasing with t. In an uninformative world, the marginal effect is zero.*

Proof.

In an informative world,

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

$$\frac{\partial w_{ijt+1}}{\partial \bar{r}_{it}} = \frac{\partial\, \mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it}]}{\partial \bar{r}_{it}}$$

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

$$\mathbb{E}[\theta_i | x_i, r_{i1}, ..., r_{it}] = \alpha + \beta x_i + \lambda_1 r_{i1} + ... + \lambda_t r_{it} = \alpha + \beta x_i + \frac{t(\lambda_1 r_{i1} + ... + \lambda_t r_{it})}{r_{i1} + ... + r_{it}} \bar{r}_{it}$$

so

$$\frac{\partial w_{ijt+1}}{\partial \bar{r}_{it}} = \frac{t(\lambda_1 r_{i1} + ... + \lambda_t r_{it})}{r_{i1} + ... + r_{it}}$$

where λ_τ is the regression coefficient of $r_{i\tau}$ on θ_i. Since these are all increasing in $Cov(\theta_i, r_{i\tau})$ and

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

therefore

$$\frac{\partial w_{ijt+1}}{\partial \bar{r}_{it}} > 0$$

and

$$\frac{\partial^2 w_{it+1}}{\partial \bar{r}_{it} \partial t} = \frac{(\lambda_1 r_{i1} + ... + \lambda_t r_{it})}{r_{i1} + ... + r_{it}} > 0$$

In an uninformative world,

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

$$\frac{\partial w_{ijt+1}}{\partial \bar{r}_{it}} = 0$$

$$\frac{\partial^2 w_{ijt+1}}{\partial \bar{r}_{it} \partial t} = 0$$

$\square$

A.6 Proof of Proposition 4

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,

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

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

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

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

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,

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

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

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

$\square$

A.7 Proof of Proposition 5

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,

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

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

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

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

and

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

so

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

where β^τ 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,

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

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

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

$\square$

A.8 Wage Results Robustness Checks

To test the robustness of the choice of cut point, I rerun Equation 2.7 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.

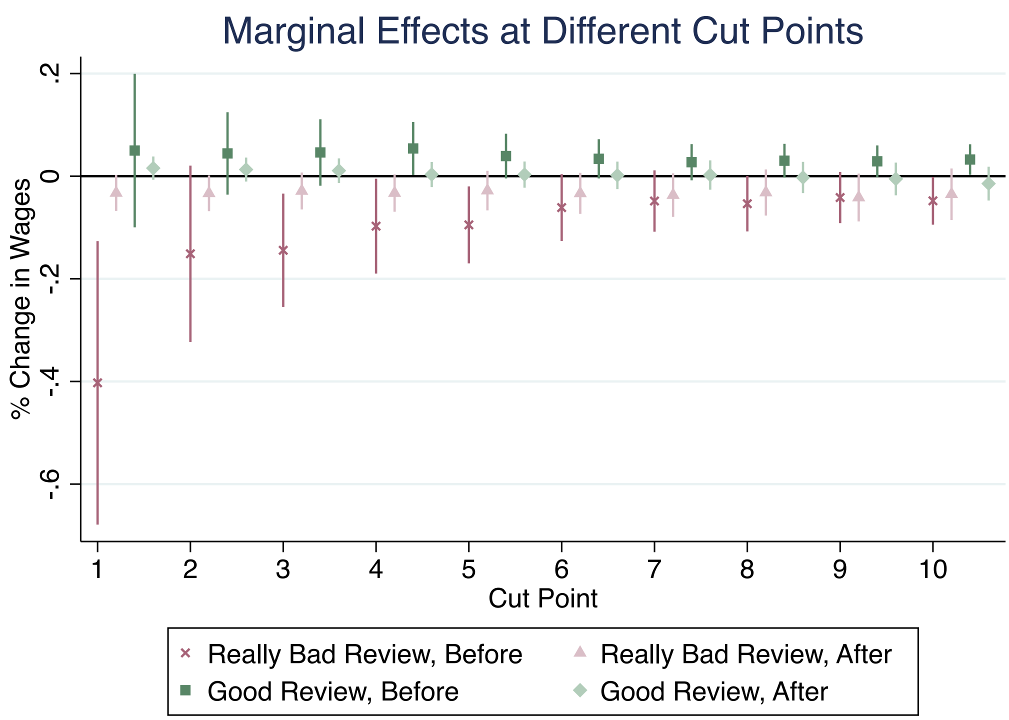


Figure A.1: Different Cut Points

I also estimate Equation 2.7 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.

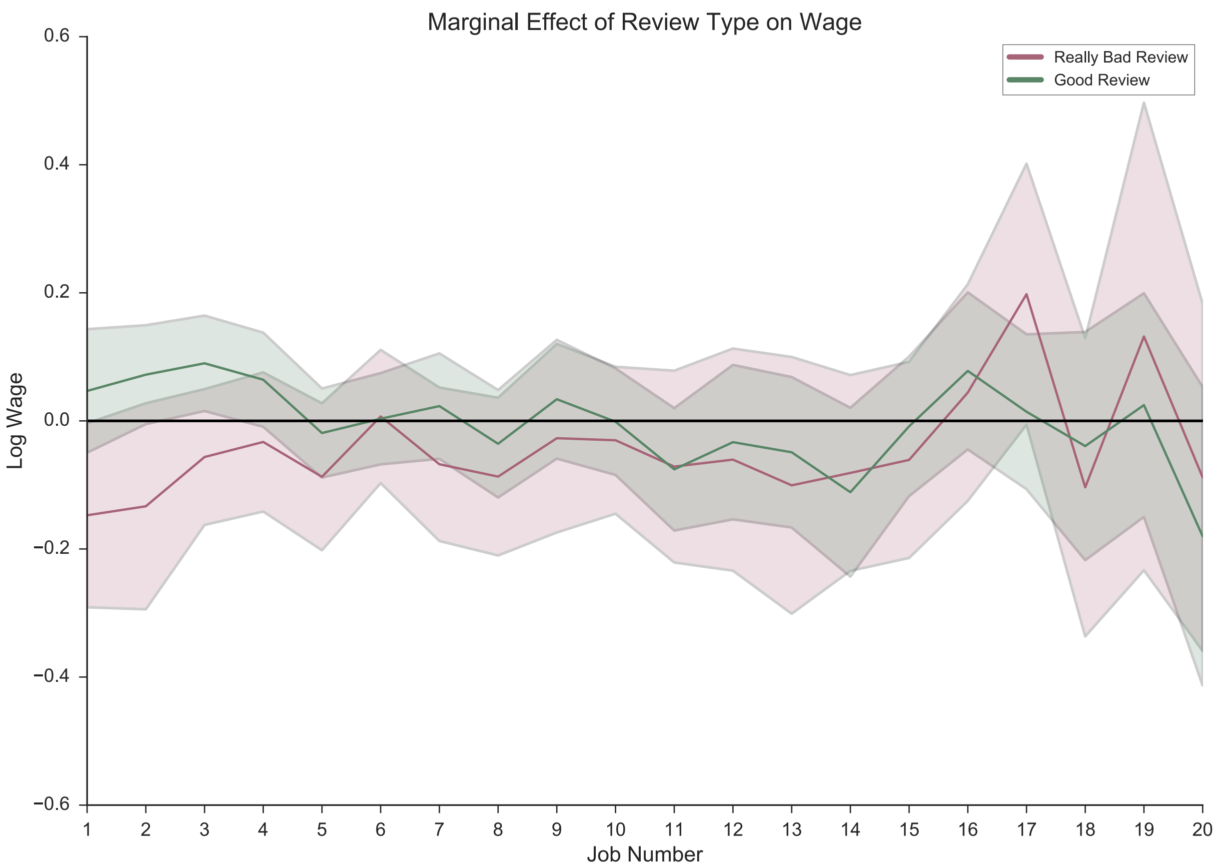


Figure A.2: Non-Parametric

A.9 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:

$$TimeToJob_{ijt} = \boldsymbol{\beta ReviewType_{it}} + \boldsymbol{\lambda X_{it}} + \boldsymbol{\mu J_{it}} + \nu_i + \kappa_j + \epsilon_{it} \tag{A.12}$$

where $TimeToJob_{ijt}$ is the time, in days, between the worker's previous job and their new one. $\boldsymbol{ReviewType_{it}}$ are indicators for the type of the most recent review, $\boldsymbol{X_{it}}$ are worker characteristics, such as education, exam scores, and previous review scores and $\boldsymbol{J_{it}}$ 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 A.2 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 A.2: Effect of Reviews on the Time to Next Job

	(1)	(2)	(3)	(4)
Really Bad Review	7.445**	5.712**	5.690**	5.268**
	(1.134)	(1.058)	(1.595)	(1.519)
No Review	3.280**	-0.579	2.180*	0.0529
	(0.678)	(0.645)	(1.054)	(1.039)
Good Review	-4.312**	-3.371**	-5.641**	-4.733**
	(0.605)	(0.569)	(0.935)	(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.9.1 Contract Renewal Results

Table A.3: Effect of Reviews on the Probability of Matching

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

Bibliography

Abowd, J., Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte. 2004. "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States." *Working Paper*.

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

Abowd, John M., Francis Kramarz, Sébastien Pérez-Duarte, and Ian M. Schmutte. 2014. "Sorting between and within industries: A testable model of assortative matching." *NBER Working Paper*.

Abowd, John M., Paul Lengermann, and Kevin L. McKinney. 2002. "The measurement of human capital in the US economy." *Technical paper TP-2002-09, LEHD, US Census Bureau*.

Abowd, John, Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." *Technical Paper TP-2002-06*.

Abraham, Katharine G., and Susan K. Taylor. 1996. "Firms' use of outside contractors: Theory and evidence." *Journal of Labor Economics*, 14(3): 394–424.

Addison, John T., Pedro Portugal, and José Varejão. 2014. "Labor demand research: Toward a better match between better theory and better data." *Labour Economics*, 30: 4–11.

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.

American Staffing Association. 2017. "Staffing Industry Statistics."

Anderson, T. W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics*, 20(1): 46–63.

Andrews, Martyn J., Len Gill, Thorsten Schank, and Richard Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3): 673–697.

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone. 2016. "Identification and Estimation of Network Models with Heterogeneous Externalities."

Avery, Christopher, Paul Resnick, and Richard Zeckhauser. 1999. "The Market for Evaluations." *American Economic Review*, 89(3): 564–584.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. "Detecting and Assessing Collinearity." In *Regression Diagnostics*. 85–191. John Wiley & Sons, Inc.

Benson, Alan, Aaron J. Sojourner, Akhmed Umyarov, and others. 2015. "Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market."

Bessen, James. 2014. "Employers aren't just whining: The 'skills gap'is real." *Harvard Business Review*, 25.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research*, 3(Jan): 993–1022.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. 2011. "Chapter 18 - Identification of Social Interactions." In *Handbook of Social Economics*. Vol. 1, , ed. Alberto Bisin and Matthew O. Jackson Jess Benhabib, 853–964. North-Holland.

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

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of Econometrics*, 150(1): 41–55.

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.

Bureau of Labor Statistics. 2017. "Establishment Data."

Cabral, Luís, and Ali Hortaçsu. 2010. "The Dynamics of Seller Reputation: Evidence from Ebay." *The Journal of Industrial Economics*, 58(1): 54–78.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. "Peer Effects and Social Networks in Education." *Review of Economic Studies*, 76(4): 1239–1267.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *The Quarterly Journal of Economics*, 131(2): 633–686.

Card, David, Ana Rute Cardoso, Jörg Heining, and Patrick Kline. 2016. "Firms and labor market inequality: Evidence and some theory." *NBER Working Paper*.

Card, David, Jörg Heining, and Patrick Kline. 2012. "Workplace Heterogeneity And The Rise Of West German Wage Inequality."

Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal. 2016. "What drives the gender wage gap? A look at the role of firm and job-title heterogeneity." *Oxford Economic Papers*, 68(2): 506–524.

Chen, Daniel L., and John J. Horton. 2016. "Research Note—Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts." *Information Systems Research*, 27(2): 403–423.

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.

Cohen, Steven, and William B. Eimicke. 2013. "Independent Contracting."

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."

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of social interactions through partially overlapping peer groups." *American Economic Journal: Applied Economics*, 241–275.

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. 2015. “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.

Fortin, B., and V. Boucher. 2015. “Some challenges in the empirics of the effects of networks.” *Oxford Handbook on the Economics of Networks, Oxford: Oxford University Press, forthcoming.*

Fortin, Bernard, and Myra Yazbeck. 2015. “Peer effects, fast food consumption and adolescent weight gain.” *Journal of Health Economics*, 42: 125–138.

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.

Gibbons, Stephen, and Henry G. Overman. 2012. “Mostly Pointless Spatial Econometrics?*.” *Journal of Regional Science*, 52(2): 172–191.

Gneezy, Uri. 2005. “Deception: The Role of Consequences.” *The American Economic Review*, 95(1): 384–394.

Goldfarb, Avi, Shane M. Greenstein, and Catherine Tucker. 2015. “Economic analysis of the digital economy.”

Gruetter, Max, and Rafael Lalive. 2009. “The importance of firms in wage determination.” *Labour Economics*, 16(2): 149–160.

Halleck Vega, Solmaria, and J. Paul Elhorst. 2015. “The Slx Model.” *Journal of Regional Science*, 55(3): 339–363.

Halliday, Timothy J., and Sally Kwak. 2012. “What is a peer? The role of network definitions in estimation of endogenous peer effects.” *Applied Economics*, 44(3): 289–302.

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*.

Hsieh, Chih-Sheng, Van Kippersluis, and Hans. 2015. "Smoking Initiation: Peers and Personality."

Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan. 2015. "Reputation & Regulations: Evidence from eBay." *Working Paper*.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. 2014. "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data." *American Economic Review*, 104(6): 1597–1629.

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." Social Science Research Network SSRN Scholarly Paper ID 1585030, Rochester, NY.

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.

Lavetti, Kurt, and Ian M. Schmutte. 2016. "Estimating Compensating Wage Differentials with Endogenous Job Mobility."

Lee, Lung-fei. 2003. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances." *Econometric Reviews*, 22(4): 307–335.

Lee, Lung-fei, and Xiaodong Liu. 2010. "Efficient Gmm Estimation Of High Order Spatial Autoregressive Models With Autoregressive Disturbances." *Econometric Theory*, 26(01): 187–230.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin. 2010. "Specification and estimation of social interaction models with network structures: Estimation of social network models." *Econometrics Journal*, 13(2): 145–176.

Lin, Xu. 2010. "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables." *Journal of Labor Economics*, 28(4): 825–860.

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).

Macis, Mario, and Fabiano Schivardi. 2016. "Exports and Wages: Rent Sharing, Workforce Composition, or Returns to Skills?" *Journal of Labor Economics*, 34(4): 945–978.

MacLeod, W. Bentley. 2003. "Optimal Contracting with Subjective Evaluation." *The American Economic Review*, 93(1): 216–240.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

Maré, David C., and Dean Hyslop. 2006. "Worker-Firm Heterogeneity and Matching: An analysis using worker and firm fixed effects estimated from LEED."

Marinescu, Ioana, and Ronald Wolthoff. 2016. "Opening the black box of the matching function: The power of words." *NBER Working Paper*.

Mariska P. 2015. "UpWork Community Forum."

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.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed representations of words and phrases and their compositionality." 3111–3119.

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.*

Mortensen, Dale T., and Christopher A. Pissarides. 1999. "Chapter 39 New developments in models of search in the labor market." *Handbook of Labor Economics*, 3: 2567–2627.

Nelson, Charles R., and Richard Startz. 1990*a*. "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One." *The Journal of Business*, 63(1): S125–S140.

Nelson, Charles R., and Richard Startz. 1990*b*. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica*, 58(4): 967–976.

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.*

Patacchini, Eleonora, Edoardo Rainone, and Yves Zenou. 2016. "Heterogeneous Peer Effects in Education." Social Science Research Network SSRN Scholarly Paper ID 2757612, Rochester, NY.

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.

Pissarides, Christopher A. 2011. "Equilibrium in the Labor Market with Search Frictions." *The American Economic Review*, 101(4): 1092–1105.

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.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature*, 43(4): 959–988.

Scheiber, Noam. 2015. "Growth in the 'Gig Economy' Fuels Work Force Anxieties." *The New York Times.*

Scott E. 2016. "UpWork Community Forum."

Sørensen, Torben, and Rune Vejlin. 2013. "The importance of worker, firm and match effects in the formation of wages." *Empirical Economics*, 45(1): 435–464.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557–586.

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).

Torres, Sónia Manuela, Pedro Portugal, John T. Addison, and Paulo Guimarães. 2013. "The sources of wage variation: a three-way high-dimensional fixed effects regression model."

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.

Wood-Doughty, Alex. 2017. "Do Employers Learn from Public, Subjective, Performance Reviews?" *Working Paper*.

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.

Yeung, Ryan, and Phuong Nguyen-Hoang. 2016. "Endogenous peer effects: Fact or fiction?" *The Journal of Educational Research*, 109(1): 37–49.

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.