

Perceived Wages versus Actual Wages: Some Facts from a Surveyed Convenience Sample

John J. Horton
Leonard N. Stern School of Business*

November 18, 2013

Abstract

Using a convenience sample of workers, I test their perceptions of hourly wages and actual hourly wages for the top one hundred BLS occupations, as measured by employment. Estimation errors are decreasing in the prevalence of an occupation. Knowing someone who has that occupation also decreases error rates. There appears to be a U-shaped pattern in knowledge about occupations. Workers do not appear to know random subsets of people with an occupation: knowledge is substantially clustered.

JEL J01, J24, J3

1 Introduction

Workers need know what various occupations pay—and will pay in the future—to plan their careers. Misinformation could lead to poorly considered human capital investments and job search strategies. However, to the extent that misinformation exists, it might be readily fixed: several studies have highlighted the powerful effects obtainable from purely informational interventions (e.g., [Jensen \(2010\)](#), [Dumas \(2009\)](#), [Card et al. \(2012\)](#)). The connection between labor market information and allocative efficiency was in part the motivation for the Knowledge of the World of Work (KWW) test, which was briefly included in the NLSY [Kohen and Breinich \(1975\)](#). However, the structure of the test made it more like an IQ test and revealed more about the test-taker than population level knowledge about jobs.

Government statistical agencies do collect and disseminate wage, educational requirements and forecasts for occupations. However, it is unclear whether this dissemination is effective. The purpose of this study is to assess the accuracy of information about occupations and explore what factors mediate this accuracy. A convenience sample of US-based workers on their knowledge about the one hundred largest (by May 2013 OES report) occupations. For each occupation, participants were asked whether (1) they knew what the occupation consisted of; (2) whether they knew anyone who held that occupation; (3) their estimate of the hourly wage for that occupation and (3) whether in the future, wages and employment for that occupation would rise or fall.

An important question is whether this convenience sample can provide insights that generalize. There are several things that suggest they would. First, [?](#) shows that the demographics of MTurk workers is similar to the US population at large. Second, we report three Google Consumer Survey results

* Author contact information, datasets and code are currently or will be available at <http://www.john-joseph-horton.com/>.

for particular occupations and find good agreement with the MTurk results. Google Consumer Surveys provides a nationally representative sample.

The advantage of using MTurk is that we can cheaply build a much larger sample than would be otherwise possible. Our main interest is in trends and patterns rather than “levels.”

For each occupation, 30 evaluations were performed. The short time taken to answer questions suggests that workers are not doing research. There was no incentive for correct answers.

The focus on the analysis was to: -characterize performance and see whether the “wisdom of crowds” hypotheses holds regarding occupations -find poorly labeled occupations in the BLS/OES -see what occupations are misperceived. -see how total employment and social knowledge mediate performance in estimating hourly wages -test whether social knowledge is “segmented” by wage

There is substantial variation across occupations in whether they are known by the sample. First, individual error rates (MSE between predicted and actual hourly wage) are decreasing the total employment in that occupation. This is both re-assuring and unsurprising. Suggesting a mechanism, the more people workers know with that occupation, the better their estimates. Knowledge of an occupation seems to be substantially wage-biased, in that workers are more accurate when predicting the hourly wage of low-wage jobs. This is true in percentages. The wisdom of crowds hypotheses suggests that combining the noisy signals of a large number of individuals can lead to very accurate predictions. This is more or less the case here: the mean worker rating for each occupation explains 70% of the variation in actual log hourly wages.

2 Results

The sample of workers answering questions on MTurk. What we lack in representativeness, we gain in numbers.

If a pattern or causal effect exists in one population, there is always a question of generalizability.

2.1 Discussion

What if the sample of respondents is just a sample of low-wage workers?

- We can ask workers for their incomes - We can show in the literature that this is not a particularly biased group - We can use Google Survey, a nationally representative sample, to adjust

What if high-wage jobs are just less common? Controlling for the total employment in that category does not make the effect go away.

What if the hourly/annual salary throws things off?

Do high-wage jobs just have more variance?

2.2 Knowledge of what different occupations are

In Figure ??, we plot the mean knowledge index by occupations.

In Figure 2 we show the fraction of respondents reporting that they know what an occupation is, binned by hourly wage. A clear U-shaped pattern is apparent: respondents know more about low- and high-wage occupations, but less about the middle. This probably reflects universal experience with low-wage jobs like waitressing and cleaning and high-wage professional jobs like doctor and lawyer.

Figure 1: Self-reported knowledge by BLS occupation title

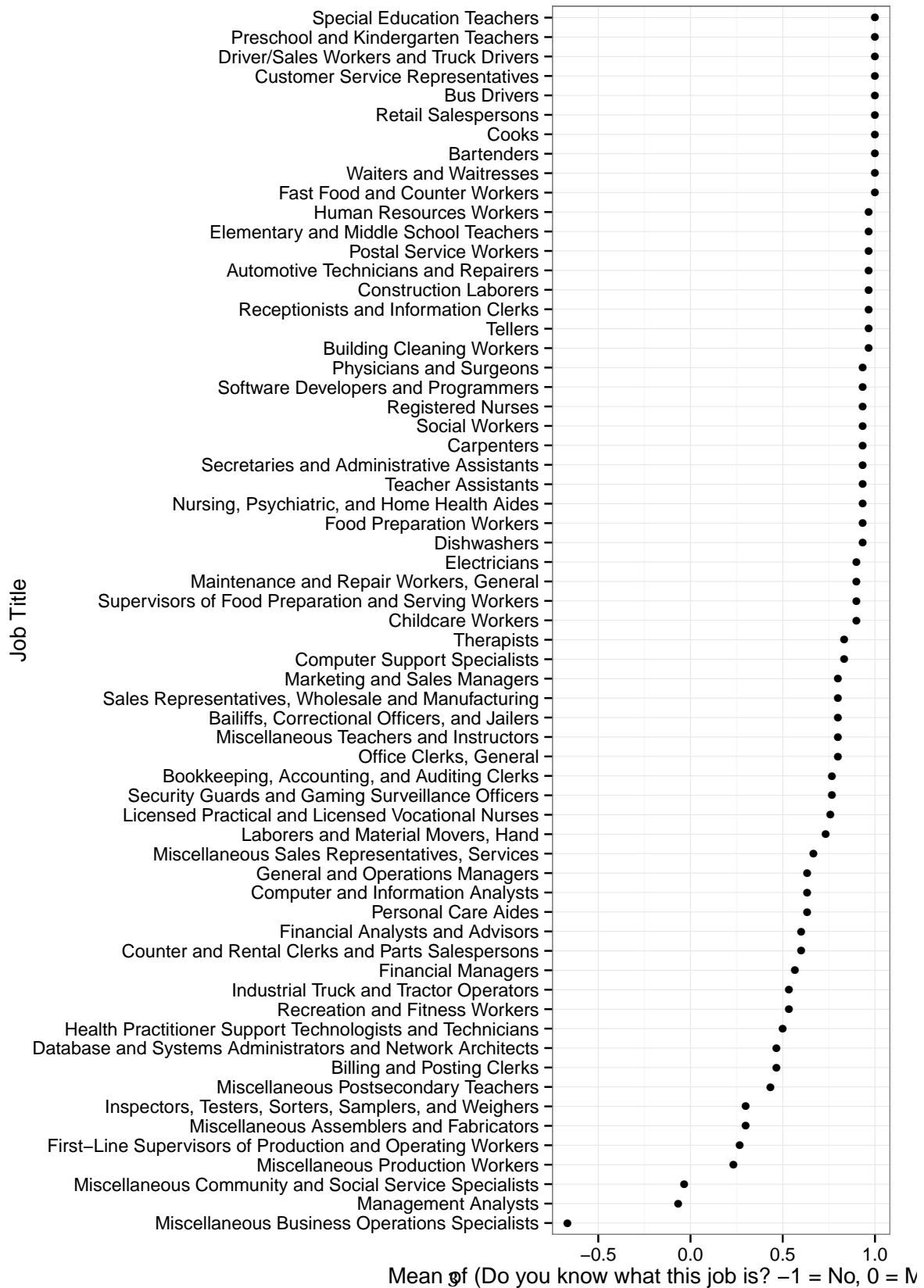
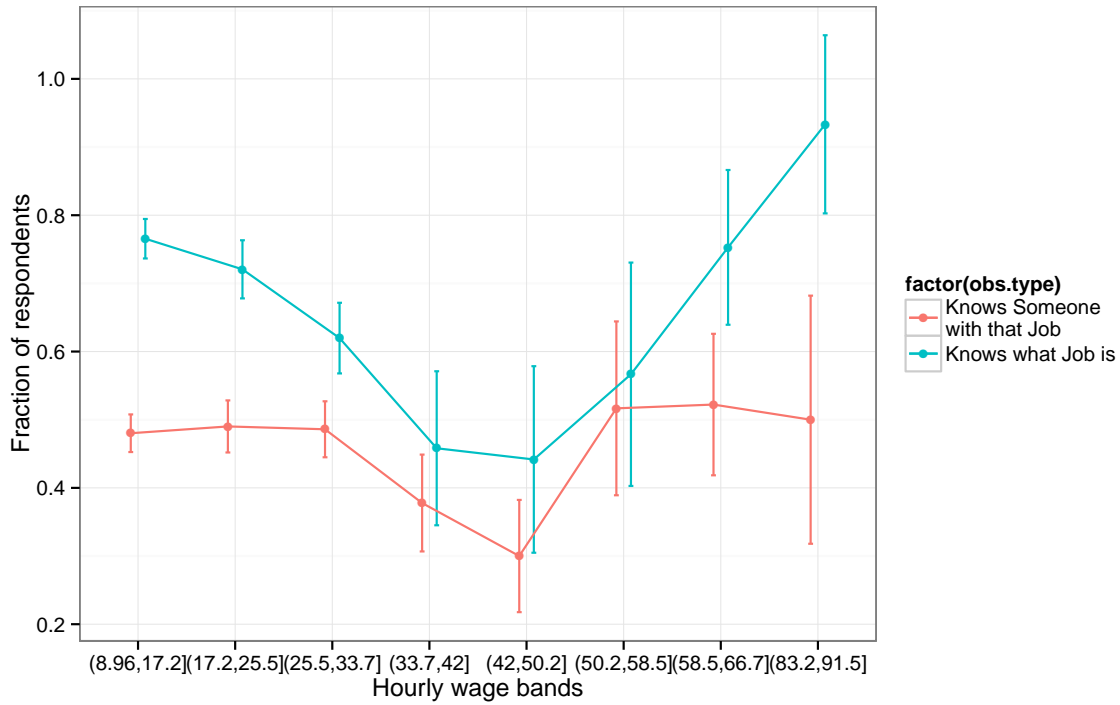


Figure 2: Hourly wage and occupational knowledge



2.3 Personal knowledge of someone working with an occupation

2.4 Prediction

In Table 1 we regress a measure of error in hourly wage estimates on a number of predictors. Column (1) reports an OLS regression, while in Columns (2) and (3) are multilevel models: Column (2) includes an occupations-specific random effect, while Column (3) includes both an occupation- and worker-specific random effect.

From Column (1), we can see that knowing what the occupation consists of is negatively correlated with the error rate, as knowing more people who have that occupation. Error is decreasing in the total employment in that occupation, though presumably both knowledge measures are mediating this total employment effect. In Column (2), when we control for occupation, we cannot include the total employment measure.

With occupation-specific controls, knowing what the occupation is still reduces error, but the effect is cut in 1/3. Much of the variation in error rates is seemingly explained by occupation-specific variation in what a job consists of. With these occupation-specific controls, social knowledge becomes slightly more effective at reducing error rates. In Column (3), the inclusion of an individual specific random effect does not lower the effect of social knowledge. This rules out the possibility that workers that know lots of people are somehow just better at predicting wage rates generically.

Figure 5 shows the mean and median wage for each occupation, along with a boxplot (without outliers) the wage predictions. The x-axis is on a log scale. We can see that individuals systematically underestimate the returns at high end of the labor market. This can be seen even more clearly in Figure 6, in which we plot mean predicted wages versus actual hourly wages.

Figure 3: Mean errors by occupation



Figure 4: Social index by BLS occupation title

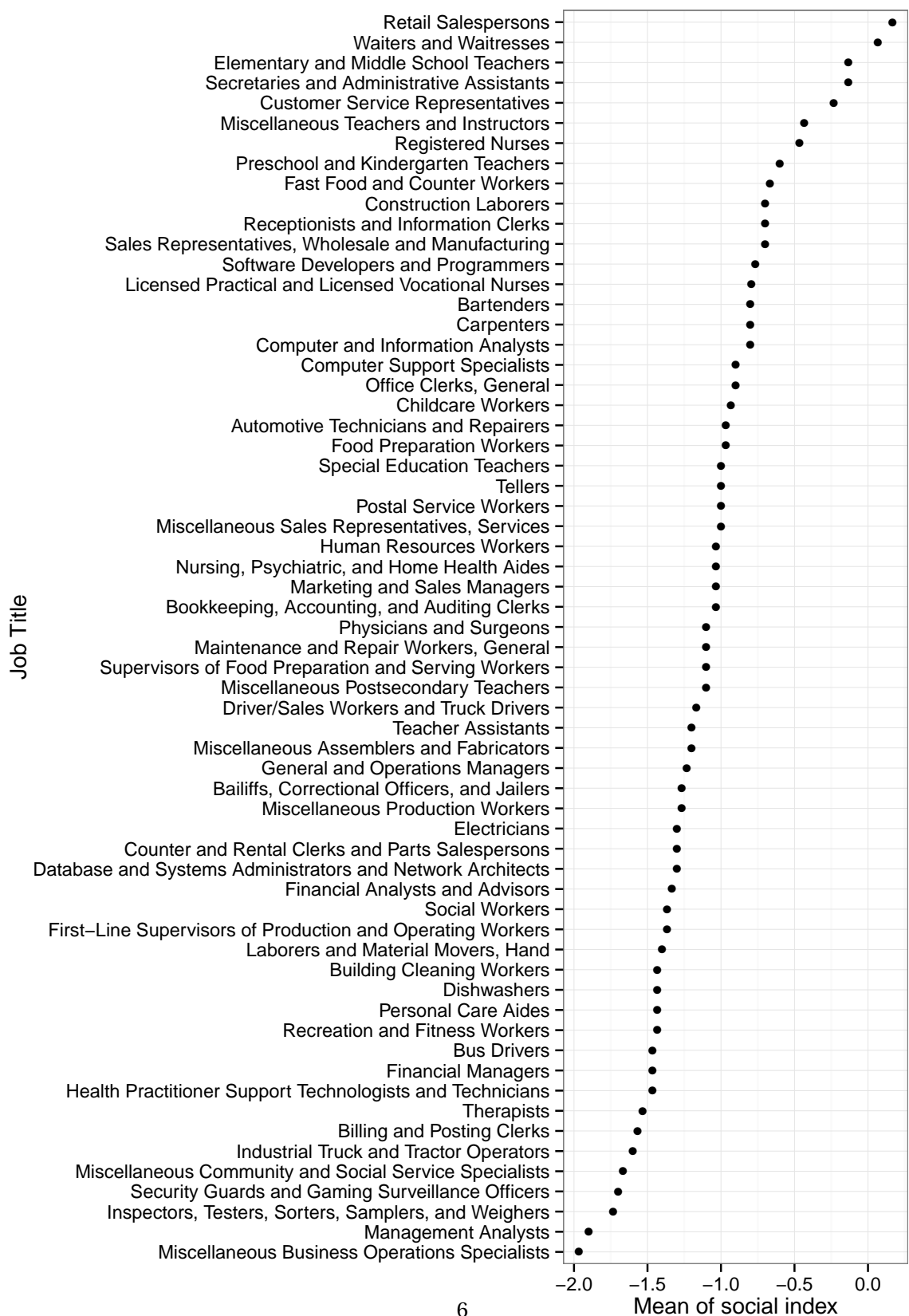


Table 1: Worker error (absolute value of difference in log wages)

	(1)	(2)	(3)
(Intercept)	0.627*** (0.110)	0.363*** (0.019)	0.369*** (0.020)
Knows someone with the job	-0.008 (0.005)	-0.011* (0.005)	-0.010* (0.005)
Knows what job consists of	-0.035*** (0.010)	-0.010 (0.009)	-0.009 (0.010)
Log total employment	-0.018* (0.008)		
Var((Intercept) title)		0.024	0.024
Var(Residual)		0.061	0.058
Var((Intercept) WorkerId)			0.004
R-squared	0.011		
adj. R-squared	0.010		
sigma	0.290		
F	10.679		
p	0.000		
Log-likelihood	-535.826	-201.886	-163.483
Deviance	248.486	403.771	326.965
AIC	1081.652	413.771	338.965
BIC	1111.603	443.723	374.907
N	2952	2952	2952

The data points are also labeled with prediction outliers. It is interesting to note that a common feature of occupations where respondents underestimate are managerial occupations. The only substantial overestimate is the wages for home health aides. This could represent a mistaken association of “health” with high-skilled (and hence high-paying).

We saw from Table TK that knowing someone who works in an occupation reduces uncertainty about that occupation. A natural question is whether workers are stratified in their knowledge. In other words, for each occupation, does a worker’s probability of knowing someone in that occupation depend upon the wages of other occupations for which they said they knew someone?

In Table 2, we explore whether respondents differ in their knowledge of someone working in a job based on the wages of people they know working in other jobs. For each observation i , we compute \bar{w}_{-i} , which is the mean wage of other occupations that worker evaluated in which they said they did know someone. We then regress:

$$\text{know}_i = \beta_0 + \beta_1 \log w_i + \beta_2 \log \bar{w}_{-i} + \beta_3 \log w_i \times \log \bar{w}_{-i} \quad (1)$$

Although high wages are associated with a lower probability of knowing what a job consists of, the interaction is positive. This suggests that respondents that reported knowing people in higher-wage occupation were more likely to know someone in the i th occupation if the i th occupation was itself higher-wage.

References

- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez**, “Inequality at Work: The Effect of Peer Salaries on Job Satisfaction,” *American Economic Review*, 2012, 102 (6), 2981–3003.
- Dupas, Pascaline**, “Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya,” Technical Report, National Bureau of Economic Research 2009.
- Jensen, Robert**, “The (perceived) returns to education and the demand for schooling,” *The Quarterly Journal of Economics*, 2010, 125 (2), 515–548.
- Kohen, Andrew I and Susan C Breinich**, “Knowledge of the world of work: A test of occupational information for young men,” *Journal of Vocational Behavior*, 1975, 6 (1), 133 – 144.

Figure 5: Box plots showing distribution of respondent hourly wage perceptions by occupation

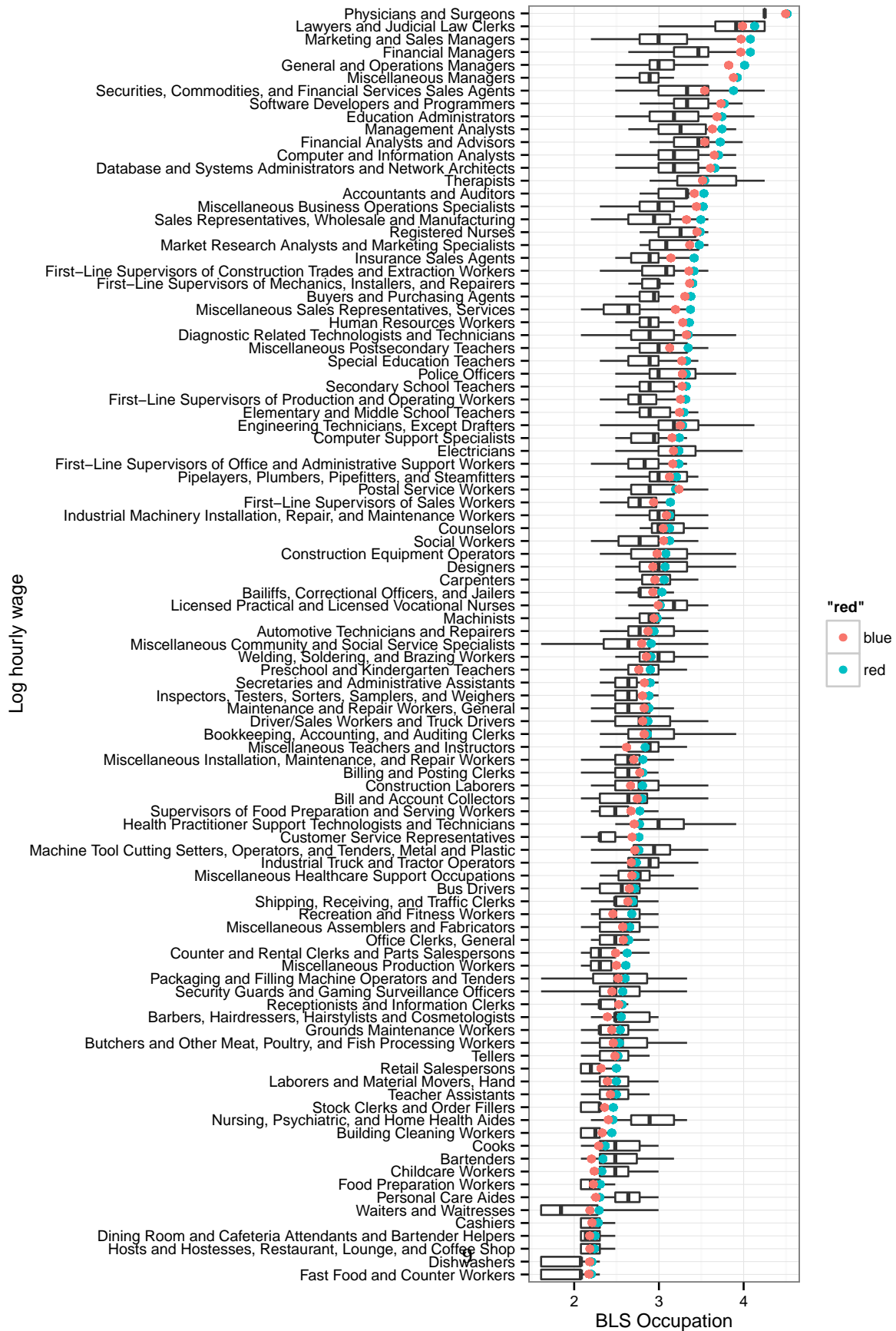


Figure 6: Perceived hourly wages versus actual hourly wages

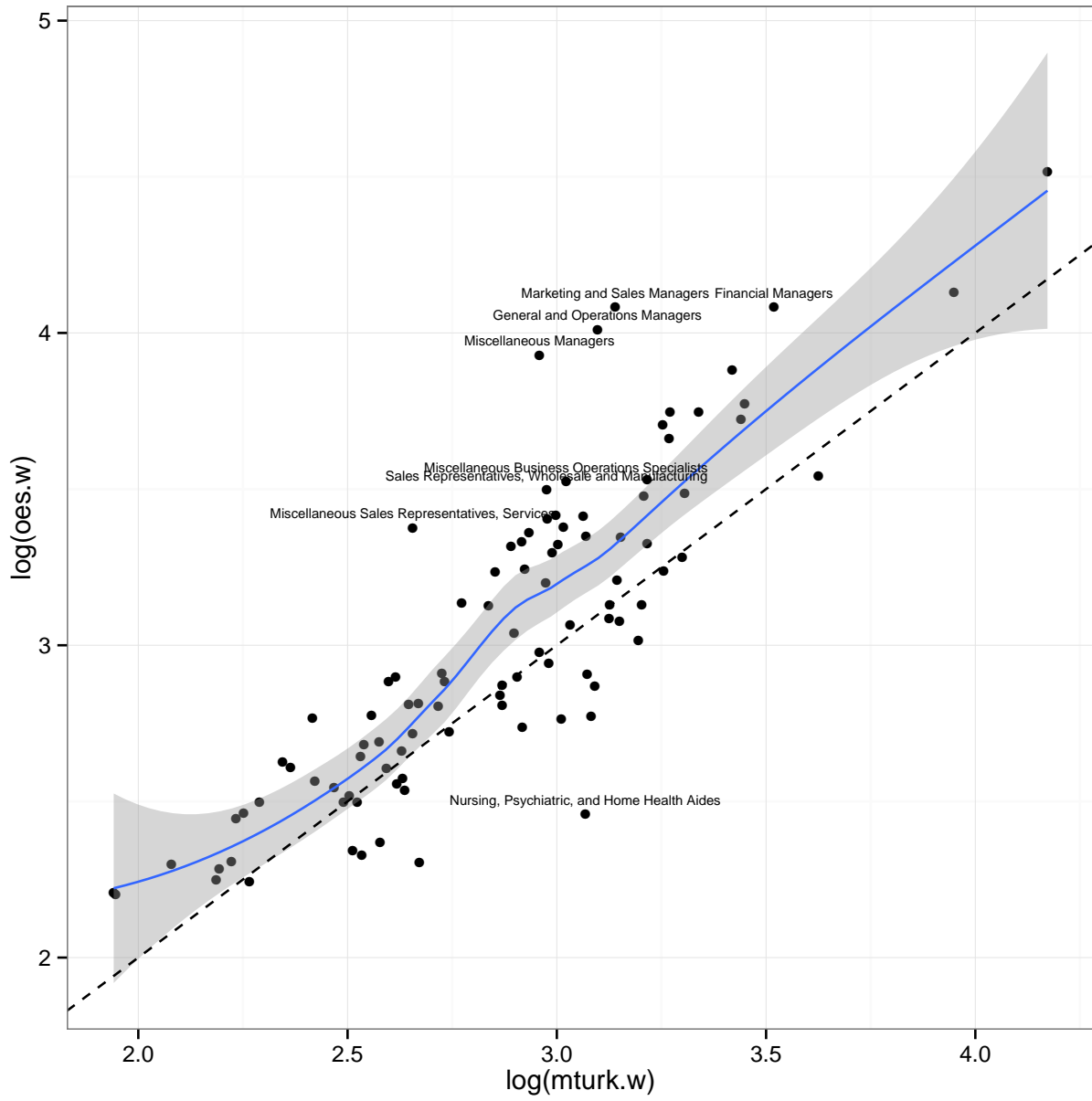


Table 2: Stratification in knowledge of someone working in an occupation

	(1)	(2)	(3)
(Intercept)	3.574*** (1.015)	2.699** (0.977)	4.880*** (0.929)
log(actual.mean.wage)	-1.510*** (0.327)	-1.244*** (0.306)	-1.390*** (0.296)
mean.wage.others.social	-1.626*** (0.332)	-1.318*** (0.320)	-1.406*** (0.307)
log(tot.emp)	0.131*** (0.015)	0.131*** (0.014)	
log(actual.mean.wage) × mean.wage.others.social	0.504*** (0.109)	0.415*** (0.102)	0.447*** (0.098)
Var((Intercept) WorkerId)		0.045	0.040
Var(Residual)		0.198	0.179
Var((Intercept) title)			0.027
R-squared	0.040		
adj. R-squared	0.039		
sigma	0.489		
F	26.252		
p	0.000		
Log-likelihood	-1754.623	-1609.371	-1555.602
Deviance	596.067	3218.743	3111.204
AIC	3521.245	3232.743	3125.204
BIC	3556.182	3273.503	3165.964
N	2497	2497	2497

Figure 7: Wage trends

