

Perceptions of Occupations: Good on Wages, Bad on Changes

John J. Horton
Leonard N. Stern School of Business
New York University*

Adam Kapelner
Department of Mathematics
Queens College, CUNY

March 12, 2015

Abstract

A sample of labor market participants were asked to estimate the wage and projected employment growth for a selection of occupations; they also reported how many people they knew working in that occupation, which was used to create a “social knowledge” index. Collectively, respondents have good knowledge about wages — their wage estimates can explain more than 50% of the variation in actual log wages as reported by the United States Bureau of Labor Statistics. However, it appears they know very little about the direction of projected employment trends; their predictions about future employment are no better than chance. Despite poor overall predictions about employment trends, individual predictions were substantially improved if the respondent had better social knowledge about that occupation.

JEL J01, J24, J3

1 Introduction

When planning a career, a labor market participant should know both what a career can pay and the future prospects of that career. Misinformation about the current and future prospects of different occupations could lead to poorly considered career and human capital allocation decisions. This paper attempts to assess the quality of labor market knowledge. To do so, we employ a convenience sample of the US population drawn from Amazon Mechanical Turk (MTurk).

*Author contact information, datasets and code are currently or will be available at <http://www.john-joseph-horton.com/>. Thanks to Alex Gedranovich for excellent research assistance.

For the top 100 most common American careers, we compare respondent per-occupation estimates of wages and respondent employment growth estimates to gold-standard estimates provided by the Bureau of Labor Statistics (BLS). The paper also attempts to understand how predictive accuracy is affected by “social knowledge” of that occupation, where social knowledge is an index based on how many people the respondent knows working in that occupation.

Collectively, the “wisdom of crowds” hypothesis seems to hold regarding wages: excluding some notable exceptions, the mean of estimates of wages were quite accurate. However, respondents did underestimate the wages of higher wage jobs, though it is unclear how much of this is due to fact that our sample is relatively poorer and younger compared to the population at large.

However, the same wisdom of the crowds does not apply when estimating future career projections. Respondents were asked whether the fraction of people employed in an occupation will go up, go down or stay the same. Collectively, they do not predict better than chance when their answers are compared to the BLS projections. However, the more social knowledge a respondent has about an occupation, the better their prediction on the employment trend for that occupation: all else equal, a 1 standard deviation increase in our social knowledge index is associated with TK.

The improvement in predictive accuracy from social knowledge is found despite our controlling for both the respondent and the occupation. This analysis rules out the possibility that some occupations are easier to predict and many people know workers in those occupations; they also rule out the possibility that people with more social connections are better predictors. While still not causal, explanations for why social knowledge seems to improve predictive accuracy seem less likely than the simplest one, namely that individual learn more about occupations when they know someone working in that occupation.

The ability to include both kinds of controls illustrates an offsetting advantage to using MTurk: while still a convenience sample, MTurk provides a way to obtain a large number of

responses even for a questionnaire that a “representative” sample would find tedious and be unlikely to complete carefully. This feature of MTurk is increasingly being used for questions in economics ([Kuziemko et al., Forthcoming](#); [Saez and Stantcheva, 2013](#)).

There is also evidence that suggests MTurk convenience samples are better than in-person convenience samples (a common sample published in social science) and can be similar to true probability samples. [Berinsky et al. \(2012\)](#) found MTurk respondents are similar in demographic profile to the gold-standard probability sample except they are younger, more politically liberal and pay more attention to cognitive tasks. We believe that the sample being younger biases our results downwards as younger people should be less likely to predict both wages and future trends due to inexperience. We do not believe political orientation or need for cognition will have an effect on the generalizability of the associations this study reveals.

There are conceptually similar papers to this one, but most of have focused on schooling decisions. For example, [Dominitz and Manski \(1996\)](#) elicited the returns to schooling from a convenience sample of high school students and undergraduates. Another example is [Jensen \(2010\)](#), which also focuses on returns to schooling, but includes an experimental information intervention that increased later schooling. General questions about the labor market — including some questions about various occupations — were part of the National Longitudinal Survey (NLSY) for a brief period in the Knowledge of the World of Work (KWW) subtest ([Kohen and Breinich, 1975](#)), though the nature of the questions was quite dissimilar to the questions asked in this paper. This is the first paper we are aware of that focuses explicitly on occupational knowledge.

There are many potential sources of knowledge about careers: career counselors, formal government statistics and informal channels like family and friends. There is a long-standing interest in how informal channels in labor markets affect outcomes ([Rees, 1966](#); [Stigler, 1962](#)), but the very informality of these channels has made them difficult to study. Interest in these informal channels — particularly referrals — is going through a renaissance as it becomes easier

to run experiments (Pallais and Sands, 2013) and obtain high-quality administrative data (Burks et al., 2013). This paper focuses on a social “stage” manifesting earlier than referrals by assessing how social connections mediate the knowledge needed to decide which opportunities to pursue. If the patterns discovered here generalize, gaps in our knowledge might be readily fixed: several studies have highlighted the powerful effects from purely informational interventions (e.g., Dupas, 2009; Card et al., 2012). If first-hand social knowledge is an important disseminator of labor market information, this suggests that more isolated individuals are exposed to a narrow slice of occupational possibilities putting them at a labor market disadvantage, thereby contributing to inequality.

2 Data collection and description

For the top 100 US occupations by employment totals as measured by the May 2013 Occupational Employment Statistics (OES),¹ respondents from Amazon Mechanical Turk were asked (1) whether they knew what the occupation consisted of (2) how many people they knew who held the occupation, if any, (3) their estimate of the hourly wage for the occupation and (4) whether in the future employment in the occupation would rise, fall or stay the same. The actual survey interface presented to respondents is shown in Figure 2 in Appendix A. There were no incentives for correct answers and the short time most workers spent on any tasks suggests they were not performing research.

A total of 127 respondents participated in the survey. Each occupation was evaluated by 30 different respondents.² We excluded occupations that a relatively high fraction of respondents reported not knowing what the occupation consisted of: to be in the sample at all, at least 20 out of 30 had to report having knowledge of the occupation. We did this because for

¹<http://www.bls.gov/oes/>

²Respondents were allowed to answer any number of the 100 different surveys for each of the 100 occupations. The median number of surveys per respondent was 13 and the range was at minimum 1 and at maximum 99. Approximately 7% of the respondents answered more than 90 and approximately 13% answered one or two.

those occupations, it is more likely that even respondents that did claim to know what the occupation consisted of were in fact mistaken. We further restricted observations to those where the respondent knew what that occupation was. Appendix B contains descriptive statistics of respondent knowledge about all occupations originally included in the survey.

For “ground truth” on occupational wages, we simply used the estimated hourly mean wage from the OES. However, for occupations where hourly wages are not available, we computed an estimated hourly wage based on a 40 hour week for 50 weeks a year. For employment predictions, we computed the percentage change from the 2010 employment level (the fraction of the workforce in each occupation) to the the 2020 employment projections. We classified changes that were $\pm 5\%$ as “Stay the Same” (getting a value of 0) and those greater than 5% getting a value of 1 and those with a greater than 5% decrease getting a value of -1 . The actual language of our survey question was “20 years” in the future. While this is not the period over which we have official BLS predictions, we worried that with too short of a time period many respondents would choose “Stay the same” and hence the request for a 20 year prediction.

For the social knowledge measure, respondents were asked how many individuals they knew in an occupation. The choices were 0, 1, 3 – 10 and “more than 10.” To construct a social index, these bins were coded as “0”, “1”, “2” and “3”, respectively and then normalized with respect to the population of all responses across occupations.

3 Results

3.1 Predictions about wages and employment growth

To compare respondent predictions to BLS estimates, we regressed the BLS estimate on the associated respondent estimate. In Column (1) of Table 1, the outcome variable is the log of the actual mean wage for the occupation. The regressor is the log of the respondent’s wage estimate for that occupation. The regression includes a respondent-specific fixed-effect and

standard errors are clustered at the level of the respondent. We can see that the estimated wage is not only highly significant and positive, but the R^2 for the model is quite high, with these estimates collectively explaining more than half the variation, even after the degrees of freedom correction.

Table 1: Predictions about wages and employment trends

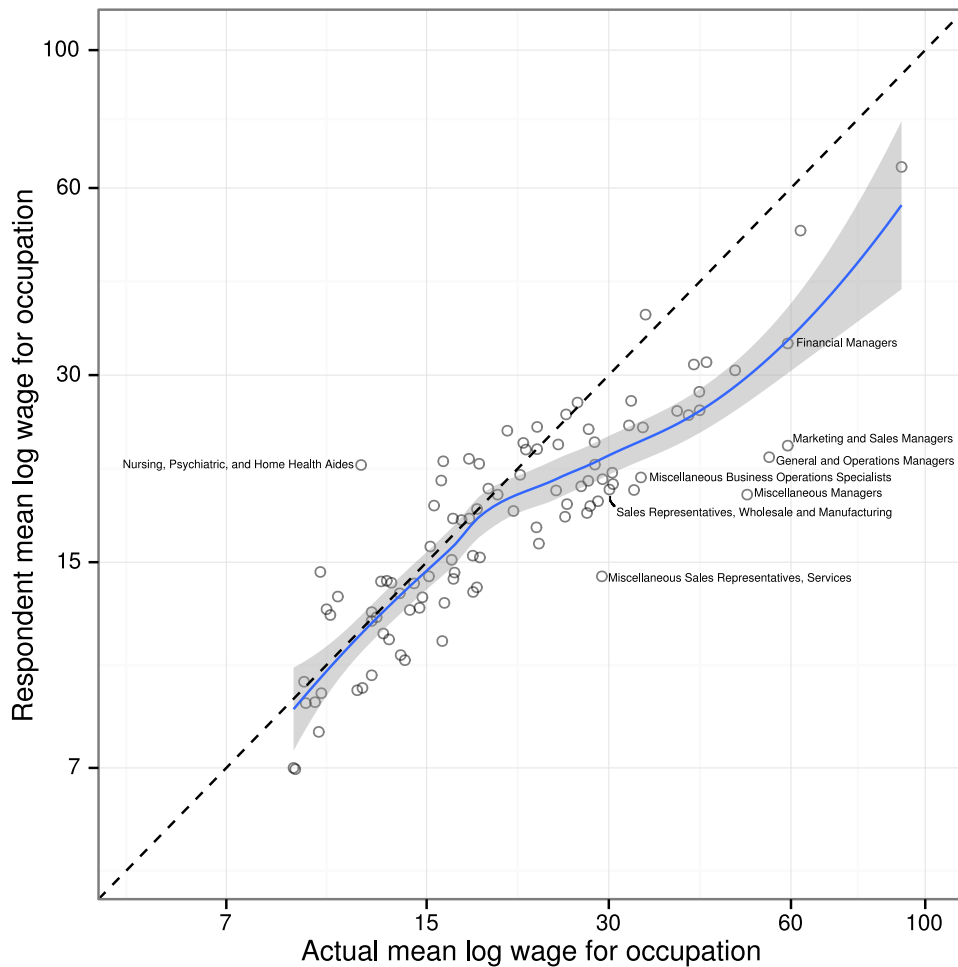
	<i>Dependent variable:</i>	
	Log Mean Occupational wage	Pred. Employment change
	(1)	(2)
Estimated log wage	0.765*** (0.023)	
Estimated employment change		-0.030* (0.013)
Respondent FE	Yes	Yes
Observations	1,898	1,898
R ²	0.569	0.059
Adjusted R ²	0.540	-0.004

Notes: The sample for both regressions reported in this table are all the responses where the respondent said they knew what an occupation consisted of. In Column (1), the dependent variable is the actual log mean wage for that occupation (from the May 2013 OES estimates from the BLS), while in Column (2) the dependent variable is an index for whether the BLS predicts that by 2020 the fraction employed in that occupation will grow by more than 5% (outcome = 1), fall by more than 5% (outcome = -1) or will not change by more than 5%, positive or negative (outcome = 0). The important independent variable in both regressions is the respondent's prediction. Both regressions include a respondent-specific standard error. Standard errors are clustered at the level of the individual respondent. Significance indicators: $p \leq 0.05$: *, $p \leq 0.01$: ** and $p \leq .001$: ***.

The regression approach masks a non-linear pattern in the relationship that is readily visible when actual wages are plotted against mean worker estimates. Figure 1 is a scatter plot of the log actual wages (y-axis) versus the mean of log predicted wages (x-axis) for each occupation. Deviations from the 45 degree line indicate prediction errors. There are two notable features of the data clearly illustrated by the the plot. We can see that respondents systematically under-

estimate the wages of occupations at the high end of the distribution.³ Some outliers — both positive and negative — are labeled with the occupation title. Many of the extreme outliers are under-estimates of the wages of managerial occupations.

Figure 1: Perceived hourly wages versus actual hourly wages



Notes: This plot shows the actual log mean wage versus the mean log wages versus predicted wages.

In Column (2), Table 1, the outcome variable is the expected change in employment on the “Go Down” = -1, “Stay the same” = 0, “Go up” = 1 scale and the independent variable is the respondent’s prediction. As with the wage predictions, a respondent-specific fixed effect is

³The only occupation that workers substantially over-estimate is the home health aide category — perhaps because of a mistaken belief that this is the nursing occupation.

included and standard errors are clustered at the respondent level. To the extent that the regression has any predictive power, it is gained by *negatively* weighting predictions: in sharp contrast to the wage prediction results, there is essentially no relationship between BLS projections and individual projections.

3.2 Social knowledge and predictive accuracy

We now consider whether social knowledge of an occupation improves individual predictive accuracy. In Table 2, Column (1), the dependent variable is the respondent's prediction error in wages, defined as the absolute value in the difference in log wages between the respondent's estimate and the BLS value. In Table 2, Column (2), the dependent variable is the respondent's prediction error in employment trends, similarly defined as the absolute value in the difference in projects. For both regressions, both occupation and respondent fixed-effects are included and standard errors are clustered at the respondent level.

Table 2: Social and errors in occupational wage and employment trajectories

	<i>Dependent variable:</i>	
	Wage Error	Employment Trend Error
	(1)	(2)
Social Knowledge Index	−0.030 (0.025)	−0.054** (0.019)
Respondent FE	Yes	Yes
Occupation FE	Yes	Yes
Observations	1,898	1,898
R ²	0.373	0.327
Adjusted R ²	0.304	0.252

Notes: This table reports the effects of social and general knowledge on a respondent's prediction error, while controlling for worker and occupation specific effects. In Columns (1) the dependent variable is MSE in wage prediction, while in Column (2), the dependent variable is an indicator for incorrectly predicting the direction of employment growth in that occupation. Significance indicators: $p \leq 0.05$: *, $p \leq 0.01$: ** and $p \leq .001$: ***.

For wages, in Column (1), we can see that social knowledge is not associated with reduced error. Thus, knowing more individuals in an occupation does not seem to correspond to knowing how much an average employee is paid in that occupation. For project employment in Column (2), however, we can see that greater social knowledge is strongly associated with lower error in projected employment. The effect size is small however and not easily interpretable due to the ordinal scales for both social knowledge and future trend.

4 Discussion

One possible explanation for error in prediction for high-wage occupations is that occupations at the high end are more likely to be salaried, which may weaken workers' ability to infer hourly wages. Also, workers on Amazon Mechanical Turk tend to have lower than average wages and tend to be younger than the population at large ([Berinsky et al., 2012](#); [Ipeirotis, 2010](#)). It is an open question whether this finding would generalize to the larger population.

The lack of predictive ability about employment trends is perhaps, *ex post*, unsurprising. Knowing employment trends for an occupation would require a fairly sophisticated understanding of long-term trends, such as the shift away from manufacturing and towards a service-based economy and the erosion of "middle skill jobs" ([Autor et al., 1998](#)). In contrast, learning the wages for an occupation is fairly straightforward.

What is surprising is that knowledge of employment trends is apparently improved through social knowledge. In terms of mechanisms, for the social knowledge effects, it seems likely that knowing someone who works in an occupation would grant access to more nuanced views of an occupation: someone working in a particular field, especially one with some tenure, could probably explain how it was changing in terms of the accomplishment of the main work task, what new technologies were being used, where new sources of supply and demand were coming from and so on.

5 Conclusion

The main result of the paper is that respondents seem to have good knowledge about wages, at least collectively, but that their knowledge of projected employment — at least as predicted by the BLS — is very poor. However, social knowledge of an occupation, i.e., knowing someone with that occupation, tends to reduce predictive error about employment trends in that occupation, even when controlling for both the respondent and the occupation. By controlling for both the individual and the occupation, we can plausibly isolate the effect of social knowledge. It is not the case — as it would be in cross-sections — that individuals with lots of social knowledge have reduced error rates or that occupations that are well-known socially also have little error. While social knowledge is of course not as good as randomly assigned, reverse causation seems implausible (i.e. accurate beliefs about a particular occupation causes one to have friends with someone in that occupation).

Given the seeming importance of social knowledge, it might be useful to expose young labor market entrants or near entrants to a wider cross section of individuals in occupations.

References

- Autor, David H, Lawrence F Katz, and Alan B Krueger**, “Computing Inequality: Have Computers Changed the Labor Market?,” *Quarterly Journal of Economics*, 1998, pp. 1169–1213.
- Berinsky, A J, G A. Huber, and G S Lenz**, “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk,” *Political Analysis*, 2012, 20 (3), 351–368.
- Burks, Stephen, Bo Cowgill, Mitchell Hoffman, and Michael Housman**, “The Value of Hiring through Referrals,” *Available at SSRN 2253738*, 2013.

- 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.
- Dominitz, Jeff and Charles F. Manski**, “Eliciting Student Expectations of the Returns to Schooling,” *The Journal of Human Resources*, 1996, 31 (1), pp. 1–26.
- Dupas, Pascaline**, “Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya,” Technical Report, National Bureau of Economic Research 2009.
- Ipeirotis, Panagiotis G**, *Demographics of Mechanical Turk* number NYU Working Paper, CEDER-10-01, NYU Working Paper Series, 2010.
- 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.
- Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva**, “How elastic are preferences for redistribution? Evidence from randomized survey experiments,” Technical Report Forthcoming.
- Pallais, Amanda and Emily Glassberg Sands**, “Why the Referential Treatment? Evidence from Field Experiments on Referrals,” Technical Report, Working Paper 2013.
- Rees, Albert**, “Information networks in labor markets,” *The American Economic Review*, 1966, 56 (1/2), 559–566.
- Saez, Emmanuel and Stefanie Stantcheva**, “Generalized social marginal welfare weights for optimal tax theory,” Technical Report, National Bureau of Economic Research 2013.

Stigler, George J, “Information in the labor market,” *The Journal of Political Economy*, 1962, 70 (5), 94–105.

Figure 2: Survey questions

Answer Questions About Occupations

The job is: \${Title}

Do you know what doing this job consists of?

- ☐ Yes
- ☐ Maybe
- ☐ No

How many people do you know personally that do this job (if any)?

- ☐ 0
- ☐ 1
- ☐ 2
- ☐ 3 to 10
- ☐ More than 10

How much do you think this job pays per hour in the US, on average?

\$7.00/hour

In the next 20 years, do you predict that the fraction of the working population doing this job will:

- ☐ Go down
- ☐ Stay the same
- ☐ Go up

In the next 20 years, do you predict wages for this job will:

- ☐ Go down
- ☐ Stay the same
- ☐ Go up

Any comments or questions?

Notes: Unfortunately there was a typo in the survey, though presumably very few if any subjects misunderstood what we hoped to convey.

A Survey Materials

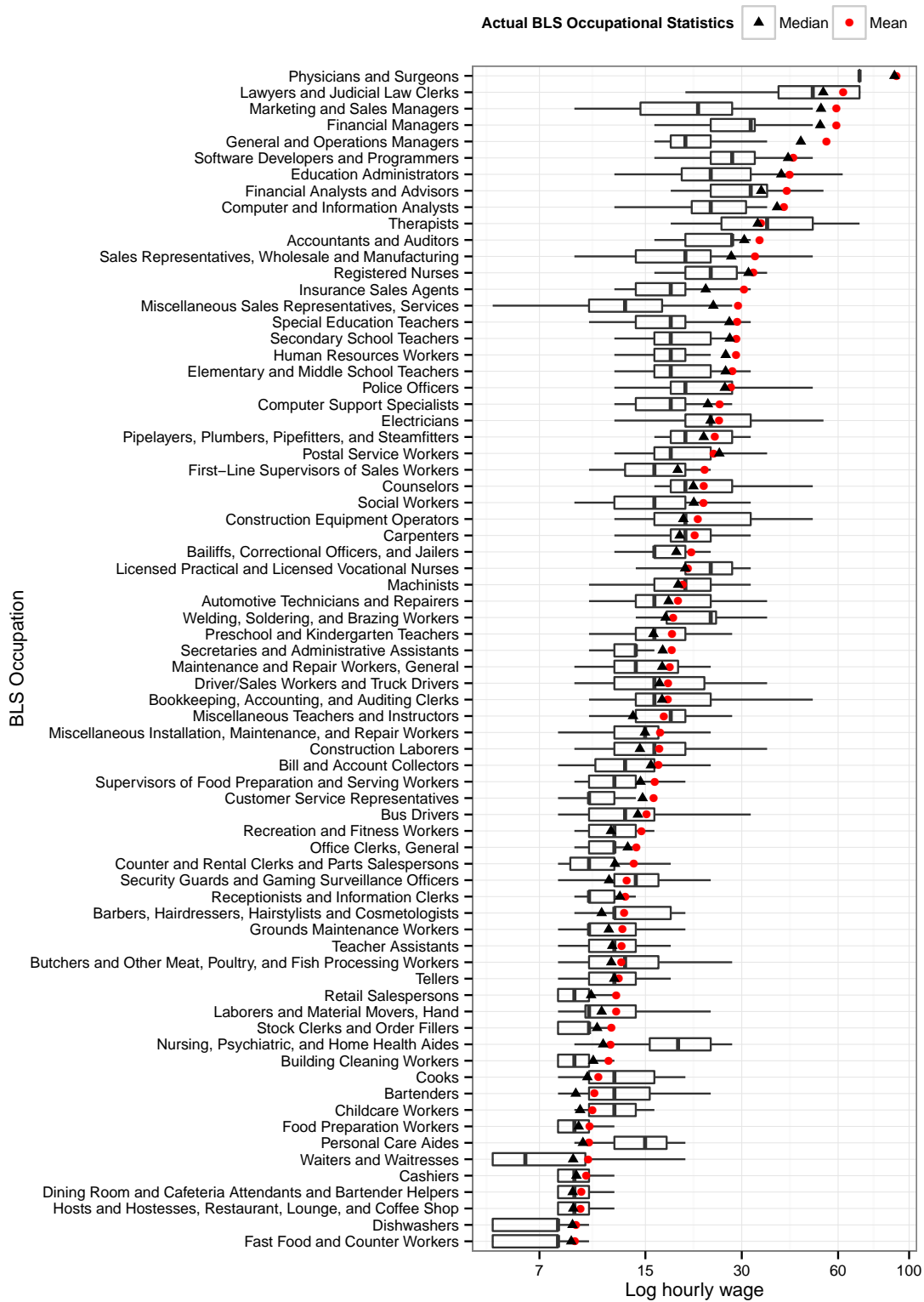
Figure 2 shows the actual survey task. Wages in the drop-down box were listed in dollar increments from \$7 to \$10, two dollar increments from \$10 to \$20, four dollar increments from \$24 to \$36 and then six dollar increments to \$70/hour.

B Additional data on occupational knowledge

Figure 3 we plot the distribution of individual respondent log wage estimates for each occupation as a box plot. The actual mean and median wage for each occupation are shown as squares and triangles, respectively. Occupations are ordered by mean wage, from highest to lowest. By comparing the median of the box plot to the actual median, we can see the tendency of respondents to systematically underestimate the wages at the high end of the distribution.

In Figure 4, we plot the mean knowledge index by occupations.

Figure 3: Distributions of respondent hourly wage perceptions by occupation



Notes: Boxplots of the respondent wage predictions are shown for each occupation, with actual mean and median over-layed.¹⁵

Figure 4: General Knowledge Index by BLS occupation title

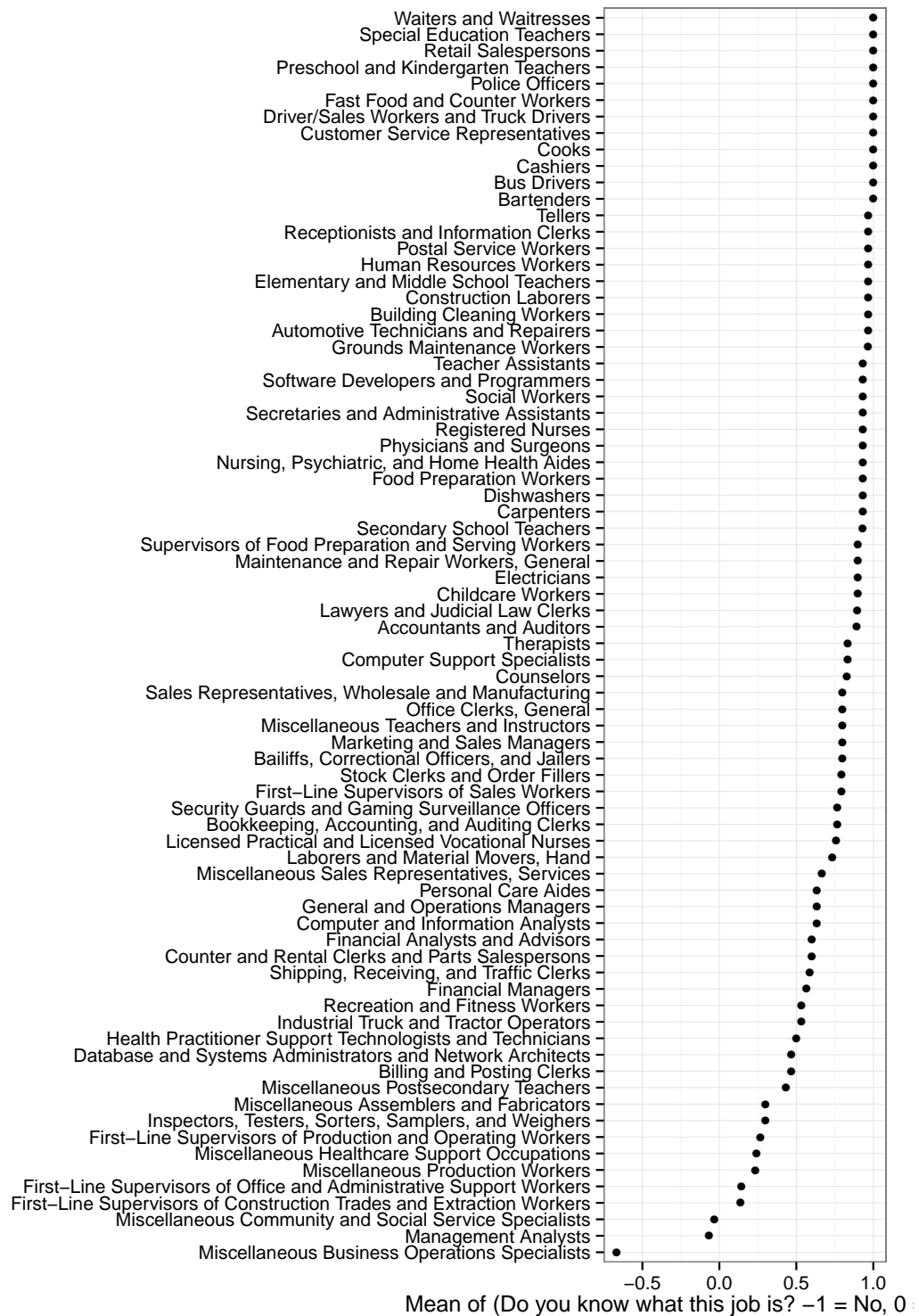


Figure 5: Social Knowledge Index by BLS occupation title

