What Labor Supply Elasticities do Employers Face? Evidence from Field Experiments*

Claus C Pörtner
Department of Economics
Albers School of Business and Economics
Seattle University, P.O. Box 222000
Seattle, WA 98122
cportner@seattleu.edu
www.clausportner.com
&
Center for Studies in Demography and Ecology
University of Washington

Nail Hassairi
Department of Economics
University of Washington
hassairi@uw.edu

December 2018

*We would like to thank Yoram Barzel, V. Joseph Hotz, Todd Sørensen, Patrick Turner, and participants at the European Society for Population Economics' annual meeting, the Pacific Northwest Labor Workshop, the Society of Labor Economists' annual meeting, and the Southern Economic Association's annual meeting for helpful comments. The research was conducted under IRB approval from University of Washington and Seattle University. The research is financed by National Science Foundation award number 1110965, "Interaction Economics: Instruments that Measure Social-Computational Systems." Partial support for this research came from a Eunice Kennedy Shriver National Institute of Child Health and Human Development research infrastructure grant, 5R24HD042828, to the Center for Studies in Demography and Ecology at the University of Washington. This study is registered in the AEA RCT Registry and the unique identifying number is: "AEARCTR-0003541".

Abstract

We provide experimental evidence on the labor supply elasticity faced by employers, which is an essential measure of employer market power. We offered two different types of jobs, each with large randomized variations in pay, and observed the amount of work performed. We find no evidence of the strong employer market power suggested by prior research, with our elasticities close to unity. Furthermore, elasticities based on the total amount of work are significantly larger than if we use worker-level data as prior studies have done. Finally, elasticities differ by job type, suggesting that worker characteristics play a crucial role.

JEL codes: J2, J3, J42, C93

1 Introduction

In the standard model of perfectly competitive labor markets, employers face an infinitely elastic labor supply. Estimates of the elasticity of the labor supply curve faced by employers are, nevertheless, mostly small, suggesting that employers have substantial market power (Manning, 2011). Many of the estimates are, however, from the public sector and obtaining unbiased estimates of the elasticity of labor supply has been complicated by the difficulty of separating demand and supply responses and by the endogeneity of observed wages and labor supply decisions.

To address these concerns, we provide experimental evidence on the elasticity of the labor supply curve faced by employers in an unregulated labor market with many employers and workers, free entry and exit, little or no discrimination, and flexible work arrangements. We run field experiments that allow us to directly estimate the labor supply elasticities that a for-profit employer face. Contrary to prior research, we can estimate not just the recruitment elasticities but also the elasticity of the total amount of work supplied.

We find employer-facing labor supply elasticities close to one, depending on job type and specification. These elasticities suggest little employer market power, especially given the short-run nature of our experiments. Furthermore, the standard in the literature is to use worker-level data to estimate the elasticity faced by employers, but we show that elasticities based on the total amount of work provided at employer-level are substantially higher than those based on worker-level data. Finally, the elasticities differ across jobs, suggesting that worker characteristics such as skills and preferences play a crucial role in determining elasticities, even within the same labor market.

Our work is made possible by the emergence of online labor markets for micro-tasks. We use Amazon's Mechanical Turk (www.mturk.com), which allows us to control all aspects of the jobs we offer, including pay. Mechanical Turk is of particular interest because it appears to be as close to a standard neoclassical labor market as one can get in a developed country setting. There is minimal regulation of the market, with the two only requirements being that a worker has to be 18 years or older and, if in the US, that the worker has a bank account and social security number.[1] Furthermore, it is a piece-rate labor market, where workers decide from task to task whether to continue to work and if they find a given job unattractive they move on to another job without occurring any penalties.

We offered two separate jobs on Mechanical Turk: One asked workers to tag images with keywords, and the other to write letters. Each job required different skills and appealed to workers with different interests. We chose these jobs to mimic as closely as possible the other jobs available in the labor market. In each job, we randomly allocated arriving workers to different wages and observed workers' decision on whether to work or not and the amount of work supplied if the job is accepted.

Most of the extensive literature on labor supply focuses on individuals' trade-off between leisure and work, captured by the total number of hours worked.[2] A smaller number of papers, either explicitly or de facto, estimate the elasticity of labor supply faced by an employer—although researchers sometimes interpret the results as estimates of individuals' labor supply elasticities.

An early example, using observational data, is the analysis of the participation decision of vendors at a single baseball stadium (Oettinger, 1999). The 127 vendors in the sample decide from game to game whether to work. The estimated elasticity mostly fall between 0.55 and 0.65. Although Oettinger frames his results as daily labor supply of vendors, the

[1] Amazon's disclaimer for Mechanical Turk reads: "The Site is a venue for Requesters and Workers to conduct transactions. Unless we are participating on the Site as a Requester, we are not involved in the transactions, and have no control over the quality, safety, or legality of Tasks or consideration for Tasks, the ability of Workers to perform Tasks to Requesters' satisfaction, or the ability of Requesters to pay for Tasks. We are not responsible for the actions of any Requester or Worker, or performing any screening of Requesters or Workers. Your use of the Site is at your own risk."

[2] Two "stylized facts" emerge from this literature. First, estimates of labor supply elasticities based on microdata are small, especially compared to what is implied by macro models (Blundell and MaCurdy, 1999; Chetty, Friedman, Olsen, and Pistaferri, 2011; Chetty, Guren, Manoli, and Weber, 2011; Manning, 2011; Keane and Rogerson, 2015). Second, intensive margin elasticities are substantially lower than extensive margin elasticities (Heckman, 1993; Chetty, 2012). Neither stylized fact conforms to standard economic theory, which has led to an extensive literature trying to explain the findings. For a recent review see Keane (2011).

absence of information about outside activities and the short time spent working at each game imply that we should instead interpret the results as the elasticity of labor supply faced by the subcontractor that hires the vendors.

Because observational data makes it challenging to separate demand and supply responses from each other, some researchers have used quasi-experimental changes to identify elasticities. Using a temporary wage increase for teachers in northern Norway as identification leads to elasticities in the 1.0 to 1.9 range (Falch, 2010, 2011). However, using a legislated change in the wage offered by Veteran Affairs hospital to combat a shortage of nurses as identification leads to very small estimated elasticities of 0 to 0.2 (Staiger, Spetz, and Phibbs, 2010). At the opposite end of the spectrum, a mandated increase in nurse aides staffing at long-term care facilities in California resulted in inverse elasticities that were not significantly different from zero, which suggest an infinitely elastic labor supply consistent with a perfectly competitive market (Matsudaira, 2014).

More recently three papers have provided experimental evidence on the elasticity of the labor supply faced by employers. In Mexico, potential government employees were randomly offered two different wages before applying to work as community development agents, and the government is estimated to face a labor supply elasticity of around 2.15 (Dal Bó, Finan, and Rossi, 2013). A public works programs in Malawi randomized the daily wage offered for doing manual labor on agricultural development projects, which resulted in an estimated elasticity of 0.15, presumably because the experiment ran during the agricultural off-season (Goldberg, 2016). Finally, analysis of five experiments and scraping of job listing lead to an elasticity of around 0.1 in the same market as we are using (Dube, Jacobs, Naidu, and Suri, Forthcoming). The five experiments all used a "honeypot" design, where workers who have already agreed to work on one job are, depending on design, either offered a randomized wage and asked if they would like to continue to perform a fixed number of tasks or recontacted with a randomized offer for a different job. In none of these three experimental papers were workers free to decide on the amount of work to provide.

As is evident, the limited literature shows little consensus on the size of the labor supply elasticities faced by employers (Sokolova and Sorensen, 2018). Furthermore, most of the literature, for identification reasons, focused on the public sector or heavily regulated industries such as medical care. Finally, the decision analyzed has uniformly been whether somebody decides to work for an employer or not, rather than the amount of effort or hours supplied.[3]

Our experimental set-up has four main advantages compared to the prior research. First, we observe *all* potential workers who look at our offered jobs, whether they decide to work or not. Second, we allow workers to decide freely the amount of work they provide. Third, the offered wages vary more than in prior experiments, with our maximum offered wage ten times our lowest offered wage. Finally, there is minimal concern that supply or demand shifts affect wages. Workers decision on whether to work does not impact anybody else's wage because of the random assignment of wages. Since we offered only short-term jobs, there is little risk that other employers change their wage in response and no substantial role for human capital accumulation.[4]

2 Experimental Design and Data

Amazon's Mechanical Turk is the largest of the emerging micro-task markets with over 100,000 registered workers (Buhrmester, Kwang, and Gosling, 2011). Individual tasks in a job are called HITs (Human Intelligence Tasks), and work is paid per task completed rather than per hour. Examples of jobs include transcribing audio recordings into text, reviewing products, labeling images, searching for information, data entry, and answering surveys.

Workers choose jobs from a list on the website that can be sorted by criteria such as

[3] The exception is Fehr and Goette (2007), but the goal of their analysis is very different from ours.

[4] That is, workers have no incentive to accept a lower wage now to accumulate job-specific human capital that could lead to higher wage later on. See French and Stafford (2017) for estimates of labor supply elasticities with and without learning by doing.

pay per HIT and posting date.[5] Workers can preview a job before accepting it, and abort the job at any time. Once a worker completes a HIT, the program automatically serves the next HIT, and the worker decides whether to continue or not. There is no minimum number of HITs within a job, and most employers have no limits on the number of HITs that a worker can complete in a day.

Mechanical Turk requires that workers have to be 18 years or older, but otherwise there are few restrictions on participation.[6] Our only requirement is that the computer accessing our jobs must be in the US. Focusing on the US allows us to estimate consistent wage responses while achieving a sufficient sample size. It is, in principle, possible to circumvent our location restriction through the use of proxy servers, but Amazon requires that workers provide a US tax ID number if their computer appears to be in the US, which limits the usefulness of proxy servers.

We offered two separate jobs that ran at different times during 2014. In one job we offered workers an image tagging task, and in the other asked them to write short letters. The two jobs were designed to be attractive to different segments within the Mechanical Turk worker community and to require different skill sets. The experiments randomized both job characteristics and pay (Pörtner, Hassairi, and Toomim, 2015). Here we focus on the randomization of pay offered.

Data collection began as soon as a worker clicked on our offered job in the job listing. We ran both jobs on our server, which combined with the way that we hooked into Amazon's API, allowed us to see the unique worker ID assigned by Amazon for all workers that clicked on our job in the listing and to randomize their offered wage. We did not inform workers that they were part of an experiment to rule out an observer effect, where workers change behavior in response to being part of an experiment. Workers do, however, know that we monitor their output, but this monitoring is akin to what one would find in any

[5] Appendix Figure A.1 shows an example of a job listing on Mechanical Turk.
[6] Requestors can require skills and "certifications" of workers, but we do not require either.

job.[7]

Our approach has three crucial differences from previous Mechanical Turk experiments, such as those described in Dube, Jacobs, Naidu, and Suri (Forthcoming). First, in the previous experiments only workers who had already accepted a job at a low pay where used, either by offering them a randomized bonus to perform a given number of new HITs or a randomized wage for future work, also for a given number of HITs. Second, the previous experiments informed workers that they were part of an experiment and workers were asked to provide personal information. Both the reliance on prior workers and knowledge of the experiment introduce the possibility of self-selection. Finally, workers were not free to choose how much work they wanted to supply in previous experiments. As we show, our approach leads to substantially higher elasticities.

To ensure that workers who show up at different times of the day are equally likely to be presented with all job characteristics and pays, we created a list of all possible combinations in random order. Each worker that looks at our job is automatically assigned the next combination in the list, and always shown the same set of conditions based on their worker ID number after that. We observe whether the worker accepts the job and, if so, how many HITs the worker perform.

The image tagging job is straightforward and similar to many other tagging jobs offered on Mechanical Turk, where requestors have workers go through images to train a machine learning process to identify objects. Once a worker accesses the image tagging job, our program selects five pictures, and for each image, we ask the worker to provide five tags or keywords in addition to clicking a radio button indicating whether each the image is appropriate for a general audience. Figure 1 shows part of the page presented once a worker accepts the HIT, including one image. Our program randomly assigned workers to a pay per five images tagged, equal to 25 tags, of between USD 0.05 and USD 0.50. The pay varied in USD 0.05 increments.

[7] Each job had built-in automated quality checks to ensure, for example, no copying of responses or slight variations in spelling passed off as different responses.

Figure 1: Image Tagging Experiment

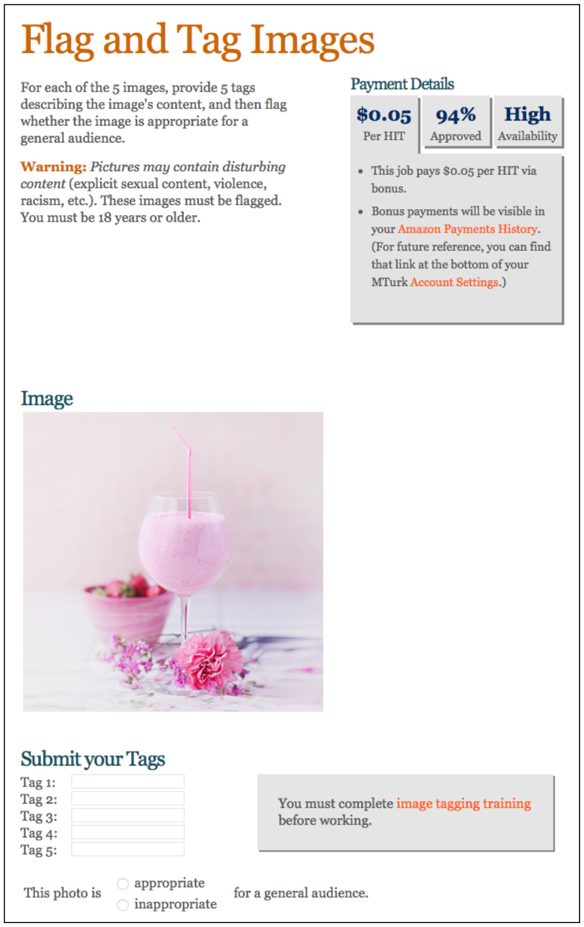
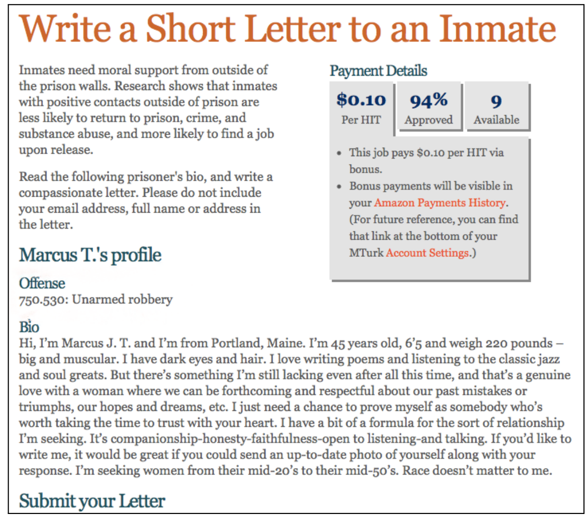


Figure 2: Letter Writing Experiment



The experiment ran over six days in 24-hour segments from 07.58 GMT.[8] A worker would see one set of conditions during a 24-hour segment and then after 07.58 GMT the worker job conditions and pay would be randomized anew. We choose 07.58 GMT because that was the time of the day where there was the fewest number of workers on Mechanical Turk. Job characteristics and pay are only random the first time a worker visits the job. For the image tagging experiment we, therefore, use only the data from the first day we observe a worker, which is not necessarily the first day the experiment ran.

In the letter writing job, the task is to write a positive and supportive letter to a prison inmate. Figure 2 shows an example. The pay varies in 10 cent increments from USD 0.1 to USD 1.0 per HIT completed, and, as in the image tagging job, this pay remained with

[8] We did not advertise the duration, as is standard on Mechanical Turk.

the worker throughout the experiment. The letter writing experiment ran for one 24 hour segment.

A total of 4,095 workers visited the image tagging experiment, and 2,111 workers visited the letter writing experiment. In both experiments, many workers looked at our offered jobs but decided not to work. For the image tagging experiment, 63 percent did not work, leaving 1,605 workers who completed one or more HITs on the first day they visited the job, which resulted in a total of 60,695 images tagged—equal to 303,475 keywords. For the letter writing experiment 73 percent did not work, leaving 578 workers who completed one or more HITs, and, who, in total, wrote 4,366 letters.

Both experiments had upper limits on the number of HITs that a worker could perform, 50 HITs for the image tagging experiment and 9 or 90, depending on the condition assigned to, for the letter writing experiment. Censoring also occurred for workers that were still working when the experiments stopped. Out of the 94 workers on the image tagging experiment with a censored number of HITs, 89 workers were at the 50 HIT limit, and the remaining five were still working when the experiment ended. For the letter writing experiment, ten workers were at the 90 HITs and 58 at the 9 HITs limits; none had a censored number of HITs because of the end of the experiment.

3 Estimation Strategy

Our experimental setup allows us to estimate the causal effects of wages on the amount of work supplied using two different approaches. The first relies on the worker-level information, similarly to what prior research has done. The second treats the employer as the level of observation, using the fact that we observe the total amount of work provided to the employer by wage level.

Because we observe all workers who look at our job—whether they reject or accept our jobs—we can directly model the worker-level selection into work and amount of work supplied. We first estimate the effect of offered wage on the decision to work for the employer:

$$1[H_i > 0] = \alpha + \beta \log(w_i) + \epsilon_i, \tag{1}$$

where $1[H_i > 0]$ is an indicator variable that takes the value 1 if worker i complete at least one HIT and 0 otherwise and w_i is offered pay per HIT for the worker.[9] We estimate the effect of wage on the decision to work using both a linear probability model and a Logit model.

One of the essential features of Mechanical Turk is that if a worker dislikes the offered set of wages and job characteristics it is easy to move on to another job. Workers, however, likely have imperfect information about the offered jobs. They may, for example, not know exactly how long it takes to complete a HIT, and therefore whether it is worthwhile to work for the offered wage. The combination of imperfect information about return and the ease of moving between jobs suggests that many workers will complete a "trial" HIT and then decide whether to continue working. We therefore also estimate the decision to work counting two or more HITs as working. That is, we estimate equation (1) using $1[H_i > 1]$ as the dependent variable.

We next turn to the intensive margin. To show what the intensive margin results would look like for regular firm labor data with no control for selection, we estimate the effects of wage on the number of HITs completed, conditional on workers completing at least one HIT:

$$H_i = \alpha + \beta \log(w_i) + \epsilon_i \text{ if } H_i > 0. \tag{2}$$

We present results from a censored regression model, with normally distributed errors, that takes into account the upper bound censoring built into the experiment. We also estimate equation (2) conditional on $H_i > 1$ to account for learning about the return to working on a job, in which case we use $H_i - 1$ as the dependent variable.

[9] The offered job characteristics are, by design, orthogonal to the offered wage and we, therefore, do not show the effects of job characteristics here. See Pörtner, Hassairi, and Toomim (2015) for those results.

Finally, we use information on all workers to estimate censored regression models, again assuming normally distributed errors, that take into account the upper bound censoring built into the experiment. The censored regression model implicitly requires two assumptions: that we observe wages for all workers independent of whether they work or not, and that wages are exogenous to the workers' labor supply (Blundell, MaCurdy, and Meghir, 2007). Neither assumption would be acceptable in standard labor market data but are appropriate here. Again, we also estimate conditional on $H_i > 1$, in which case we use $H_i - 1$ as the dependent variable.

We calculate three worker-level based wage elasticities for each experiment: the extensive margin elasticity, the intensive margin elasticity conditional on working, and the overall elasticity for all workers observed whether they accept our job or not.

The extensive margin elasticity captures the effect of wage changes on the probability of doing any work on our offered job. The coefficients from the extensive margin regressions do not directly show the extensive margin elasticity, but for the linear probability model we can calculate it as

$$\epsilon_e = \frac{\partial \Pr[H > 0] / \Pr[H > 0]}{\partial w / w} = \frac{\beta}{\Pr[H > 0]}, \tag{3}$$

using the results from our estimation of equation (1) with log wage as the explanatory variable, where $\Pr[H > 0]$ is the probability that the number of HITs performed is greater than zero.[10] For the probability, we use the proportion of workers who perform more than one HIT for each experiment. We also provide the extensive margin elasticity when counting two or more HITs as working, that is $\epsilon_e = \frac{\beta}{\Pr[H>1]}$.

Similarly, we calculate the intensive margin elasticities conditional on working as

$$\epsilon_i = \frac{\partial H / H}{\partial w / w} = \frac{\beta}{H}, \tag{4}$$

[10] For the Logit model the elasticity can be calculated as $\epsilon_e = \beta(1 - \Pr[H > 0])$.

using the results from equation (2) and with H calculated as discussed below.

From the employer's perspective, the elasticity of most interest is the overall elasticity taking into account both workers' decision to work or not and the amount of work performed. We calculate the overall elasticity using equation (4) with the estimates of β from the censored regression model using all workers.

Given the built-in censoring in the experiments, a critical question is how to calculate the average number of HITs, H. The censoring implies that if we calculate the average number of HITs as a simple average of the observed number of HITs per worker the estimated elasticities will be biased upward.[11] To correct for the censoring, we employ two different approaches, one non-parametric and one parametric. The idea behind both is that we can treat the observed number of HITs as a duration measure, where each worker either exits the job voluntarily or is censored at the built-in maximum or the end of the experiment. For each sample, we predict the average number of HITs for each wage level taking censoring into account and then average across the ten wage levels to arrive at an overall average.[12]

The non-parametric method is the Kaplan-Meier estimator (Kaplan and Meier, 1958). Estimating the mean number of HITs for a wage level requires specifying the maximum number of HITs possible. We present four different values. "None" is identical to ignoring the right-most censoring, but will account for right-censoring at less than the maximum number of HITs observed. The other three assume that the true maximum, which is the point at which the last workers would have stopped working on their own, is either 25, 50, or 75 HITs higher than the censored maximum.[13] The parametric method uses the Gamma distribution. This method does not require any assumptions about the maximum

[11] We have no such problem for the extensive margin since we observe all workers offered our jobs, whether they work or not.

[12] For the elasticities where we discard a "trial" HIT, we use the number of HITs minus 1 as the basis for these calculations.

[13] The Kaplan-Meier estimates are close to what we get if we assume a uniform distribution of the number of HITs that censored workers would perform between the built-in maximum and the assumed true maximum.

possible number of HITs but instead relies on the functional form to estimate the average number of HITs performed at each wage level.

One of the advantages of our experimental set-up over prior research is that we observe the total amount of work provided by wage level. Hence, we can treat each of the ten wage levels as an observation and directly examine how the total amount of work respond to wage changes. Because of randomization, the number of workers offered each wage varied slightly across wage levels. We, therefore, base our estimates on the average number of HITs across all workers offered a given wage level. The elasticities presented are the coefficients on $\log(wage)$ from the regression

$$\log(mean\ HITs) = \alpha + \beta \log(wage) + \epsilon, \tag{5}$$

using the ten wage levels as observations.[14] As for the worker-level estimates, we need to take into account the censoring, except here a failure to account for censoring will tend to bias downward the estimated elasticities since the likelihood of censoring tend to increase as the offered wage increases. We use the same approach as above to calculate different measures of the average HITs per wage level and run the regression on these values.

There is no direct way to calculate standard errors for any of our estimated elasticities. We, therefore, use bootstrapping to obtain standard errors for each of the elasticities. We draw from the underlying sample with replacement, rerun the relevant regressions, including the censor-corrected means, and calculate the associated elasticity. For each elasticity, we perform 1,000 draws.

[14] An alternative approach is to calculate pairwise elasticities and then take a weighted average using the percent change in wage between pairs as the weight. This approach leads to slightly higher elasticities, but bootstrapping is less well behaved.

4 Elasticities Based on Worker-level Data

Table 1 shows the estimated effects of the offered wage on labor supplied for the two experiments using worker-level data. The first two columns show the results for the extensive margin using a linear probability model and a logit model, the third column shows the intensive margin results using the sample of workers who worked, and the final column shows the results using all workers. Furthermore, the table shows the uncorrected mean of the dependent variable for the relevant sample.

The effect of increasing the wage is positive and statistically significant at the 1% level in all cases. Furthermore, except for the extensive margin linear probability model for the letter writing experiment, the estimated coefficient for log wage is higher when counting two or more HITs as working than when counting one or more HITs as working.

Table 2 shows the elasticities, calculated as described above, together with the associated bootstrapped standard errors. The extensive margin wage elasticities are 0.13 for the image tagging experiment and 0.27 for the letter writing experiment when we define working as completing one or more HITs.

The low search, entry, and exit costs for individual jobs, together with imperfect information about how attractive an offered combination of pay and job characteristics is, may lead many workers to try a job as part of their search for jobs with a sufficiently high return to effort. If workers routinely do a trial HIT before deciding whether to continue working, we should see substantial higher elasticities if we ignore the first HIT. The elasticities counting two or more HITs as working are, indeed, 0.10 percentage points higher than the pure extensive margin elasticities, which is close to a doubling for the image tagging experiment and an increase of more than one-third for the letter writing experiment.

The low extensive elasticity combined with the higher elasticity when discarding the trial HIT suggests that it takes little to have workers try a job. These two results support our conjecture that the Mechanical Turk labor market is one where workers meet little friction in moving in and out of jobs but do not have perfect information on the return to

Table 1: Effects of Wage on Labor Supplied to Employer

	Image Tagging Experiment			
	Extensive		Intensive	
	Worked = 1, not = 0		Number of HITs Performed	
	LPM	Logit	Censored[b]	
Sample	Full[a]	Full[a]	Worked	Full
	Counting One or More HITs as Working			
Log wage	0.048***	0.219***	2.840***	1.310***
	(0.011)	(0.049)	(0.519)	(0.198)
Observations	4,095	4,095	1,526	4,095
Mean of dependent variable	0.373	0.373	7.612	2.837
	Counting Two or More HITs as Working			
Log wage	0.050***	0.314***	3.965***	1.261***
	(0.009)	(0.058)	(0.850)	(0.194)
Observations	4,095	4,095	921	4,095
Mean of dependent variable	0.225	0.225	10.955	2.464

	Letter Writing Experiment			
	Extensive		Intensive	
	Worked = 1, not = 0		Number of HITs Performed	
	LPM	Logit	Censored[c]	
Sample	Full[a]	Full[a]	Worked	Full
	Counting One or More HITs as Working			
Log wage	0.073***	0.400***	4.976***	1.537***
	(0.014)	(0.076)	(1.074)	(0.276)
Observations	2,111	2,111	578	2,111
Mean of dependent variable	0.274	0.274	7.554	2.068
	Counting Two or More HITs as Working			
Log wage	0.057***	0.496***	8.223***	1.465***
	(0.011)	(0.099)	(1.912)	(0.271)
Observations	2,111	2,111	326	2,111
Mean of dependent variable	0.154	0.154	11.620	1.794

Notes. Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Effects of other job characteristics controlled for, but not shown. See Pörtner, Hassairi, and Toomim (2015) for full results.

[a] The sample consists of all workers on the first day we observed them during the experiment, whether they worked or not.

[b] The "Worked" sample consists of those who worked on the job the first day we observed them during the experiment. The "Full" sample consists of all those who were offered the job the first day we observed them during the experiment. For the top panel, the definition of work is to have performed one or more HITs. For the lower panel, the definition of work is to have performed two or more HITs. There are 94 right-censored observations.

[c] The "Worked" sample consists of those who worked on the job the first day we observed them during the experiment. The "Full" sample consists of all those who were offered the job the first day we observed them during the experiment. For the top panel, the definition of work is to have performed one or more HITs. For the lower panel, the definition of work is to have performed two or more HITs. There are 68 right-censored observations.

Table 2: Wage Elasticities of Labor Supply Based on Individual Level Information

Censoring Correction Method[a]	Image Tagging: Working defined as 1+ HITs[b]	Image Tagging: Working defined as 2+ HITs[c]	Letter Writing: Working defined as 1+ HITs[b]	Letter Writing: Working defined as 2+ HITs[c]
	Extensive Margin[d]			
None required	0.13***	0.22***	0.27***	0.37***
	(0.03)	(0.04)	(0.05)	(0.06)
	Intensive Margin, Conditional on Working[e]			
Kaplan-Meier – None	0.38***	0.37***	0.53***	0.56***
	(0.05)	(0.07)	(0.09)	(0.12)
Kaplan-Meier – +25	0.32***	0.31***	0.49***	0.52***
	(0.04)	(0.05)	(0.09)	(0.11)
Kaplan-Meier – +50	0.27***	0.26***	0.47***	0.49***
	(0.04)	(0.05)	(0.08)	(0.10)
Kaplan-Meier – +75	0.24***	0.23***	0.44***	0.46***
	(0.03)	(0.04)	(0.08)	(0.10)
Gamma distribution	0.34***	0.31***	0.57***	0.55***
	(0.04)	(0.05)	(0.09)	(0.11)
	Intensive Margin, All Workers[f]			
Kaplan-Meier – None	0.46***	0.50***	0.55***	0.58***
	(0.05)	(0.05)	(0.08)	(0.08)
Kaplan-Meier – +25	0.38***	0.41***	0.51***	0.54***
	(0.04)	(0.04)	(0.07)	(0.07)
Kaplan-Meier – +50	0.33***	0.35***	0.48***	0.50***
	(0.03)	(0.04)	(0.07)	(0.07)
Kaplan-Meier – +75	0.29***	0.30***	0.45***	0.47***
	(0.03)	(0.03)	(0.06)	(0.07)
Gamma distribution	0.41***	0.42***	0.59***	0.56***
	(0.04)	(0.04)	(0.07)	(0.08)

Note. Each elasticity calculated at the mean value of the dependent variable. Bootstrapped standard errors in parentheses based on 1,000 draws; * significant at 10%; ** significant at 5%; *** significant at 1%, based on bootstrapped CI.

[a] The Kaplan-Meier estimator is a non-parametric method that takes right-censoring into account. Estimating the mean number of HITs for a wage level requires specifying the maximum number of HITs possible. "None" is identical to ignoring the right-most censoring, but will account for right-censoring at less than the maximum number of HITs observed. The other values show the results when increasing the maximum number of HITs that we assume workers would ever perform if allowed to continue working as long as they pleased. Gamma distribution indicates a parametric survival model using the Gamma distribution that takes right-censoring into account. In both cases, estimation and prediction of the mean are calculated by wage level, and then averaged across wage levels.

[b] Based on results from Table 1 counting one or more HITs as working.

[c] Based on results from Table 1 counting two or more HITs as working.

[d] Elasticities based on Linear Probability Model results.

[e] All elasticities based on the censored regression models for the sample of workers who worked on the experiment.

[f] Based on the full sample taking right-censoring into account.

working on a given job. Hence, we interpret the low extensive margin elasticities not as an indication of employer market power as in the prior literature but as evidence of a flexible labor market with low search costs. To fully understand whether market power exists, we must instead look at the amount of work supplied.

Employers interested in the responsiveness of workers' effort to wage changes, and unable to conduct full experiments, have to rely on variation in wages across the self-selected sample of workers who have decided to work for them. To mimic this situation and examine how selection affects the estimates, we show the intensive margin elasticities conditional on working. The intensive margin elasticities are between 0.24 and 0.38 for the image tagging experiment and 0.44 and 0.57 for the letter writing experiment. Counting two or more HITs as working has little effect on the intensive margin elasticity estimate for either experiment. Hence, intensive margin elasticities for the sample of workers who worked are more than twice as large as our extensive margin elasticities—using 1+ HITs as working.

Since we have the offered wages for all workers who visited our job, independently of whether they decide to work or not, we can calculate the overall intensive margin elasticities that an employer faces using all workers who looked at our offered jobs. These intensive margin elasticities are somewhat larger than those conditional on working. The overall elasticity for the image tagging experiment counting one HIT or more as working the cut-off is between 0.29 and 0.46. The equivalent elasticity for the letter writing experiment is between 0.45 and 0.59. If we define working at completing two or more HITs, the overall elasticities are between 0.30 and 0.5 for the image tagging experiment and 0.47 and 0.58 for the letter writing experiment.

5 Elasticities Based on Total Amount of Work

Table 3 shows the estimated elasticities when we treat the employer as the level of observation and regress the total amount of work on the offered wage. As above, we correct for the censoring of HITs using the Kaplan-Meier or the Gamma distribution approaches. For the Kaplan-Meier approach, the "None" correction only corrects for censoring at less than the maximum observed number of HITs, while the others assume that the actual maximum of HITs workers would work is higher by the amount indicated. Standard errors are bootstrapped with 1,000 draws.

Table 3: Wage Elasticities Based on Total Number of HITs
Performed at Each Offered Wage Level

	Censoring Correction Method				
	Kaplan-Meier[a]				Gamma
	None	+25	+50	+75	Dist.[b]
	Image Tagging[c]				
All HITs	0.54***	0.62***	0.67***	0.72***	0.62***
	(0.07)	(0.08)	(0.09)	(0.10)	(0.08)
Discard "Trial" HIT	0.63***	0.70***	0.76***	0.80***	0.74***
	(0.09)	(0.10)	(0.11)	(0.12)	(0.10)
	Letter Writing[d]				
All HITs	0.88***	0.92***	0.95***	0.99***	0.94***
	(0.18)	(0.19)	(0.20)	(0.20)	(0.18)
Discard "Trial" HIT	0.96***	1.01***	1.04***	1.08***	1.08***
	(0.23)	(0.24)	(0.25)	(0.25)	(0.24)

Note. Our randomization process resulted in a slight variation in the number of workers offered each wage level. The estimates are, therefore, based on the mean number of HITs across all workers offered a given wage level. The elasticies presented are the coefficients on $\log(wage)$ from the regression $\log(mean\ HITs) = \beta_0 + \beta_1 \log(wage)$, using the 10 wage levels as observations. Bootstrapped standard errors in parentheses based on 1,000 draws; * significant at 10%; ** significant at 5%; *** significant at 1%, based on bootstrapped CI.

[a] The Kaplan-Meier estimator is a non-parametric method that takes right-censoring into account. Estimating the mean number of HITs for a wage level requires specifying the maximum number of HITs possible. "None" is identical to ignoring the right-most censoring, but will account for right-censoring at less than the maximum number of HITs observed. The other columns show the results when increasing the maximum number of HITs that we assume workers would ever perform if allowed to continue working as long as they pleased. Estimation and prediction of the mean are calculated by wage level.

[b] Parametric survival model using the Gamma distribution that takes right-censoring into account. Estimation and prediction of the mean are calculated by wage level.

[c] There are 94 worker with a right-censored number of HITs.

[d] There are 68 worker with a right-censored number of HITs.

No matter which correction we use the estimated elasticities based on the total amount of work done are substantially larger than the estimates based on worker-level data. For the image tagging experiment, the estimated elasticities are between 0.62 and 0.80 depending on correction method and whether we discard the first HIT performed as a trial HIT. For the letter writing experiment, the elasticities vary between 0.92 and 1.08.

Hence, the estimates based on employer-based measures are close to twice the size of those based on the worker-level information. As mentioned above, the estimates based on the total amount of work are consistent across methods, with the weighted average of pairwise elasticities even slightly higher than what we find using the regression approach. Furthermore, the elasticities based on discarding a "trial" HIT are consistently higher than elasticities based on all HITs.

6 Conclusion

The main question we address here is: what is the effect of a change in the wage an employer offers on the amount of labor supplied to that employer? The size of the response measures how much market power the employer has. We provide experimental evidence on the elasticity of the labor supply curve that an individual employer faces by acting as an employer in an online labor market. We offered two different jobs on Mechanical Turk. In each job, we offered arriving workers a randomized wage and observed whether or not they accepted the job and, if so, how much they worked. The main differences between our experiments and prior research are that we allow workers to choose how much they want to work and that we can follow all workers who look at our offered jobs.

Compared to the prior literature our elasticity estimates are large, especially given the short time we offered our jobs. The elasticity that an employer faces is close to 1, depending on the job. Hence, there is no evidence of the strong market power for employers found in prior work using the same labor market. Instead, our estimates suggest an almost even split between workers and employers.

Our extensive margin estimates are low, but we take this as evidence of low entry and exit costs for workers, which allows them to try jobs without committing to working for an extended period, combined with imperfect information on the return to working on a given job. Our extensive margin estimates are close to the headline estimate in Dube, Jacobs, Naidu, and Suri (Forthcoming). Hence, their claim of employer market power may be the results of confusing the low extensive margin estimates that arise from a flexible labor market with the response in the *total* amount of work provided to a change in the wage, which is the primary outcome of interest.

Our results also provide a cautionary note concerning estimating the labor supply elasticities that employers face using worker-level rather than employer-level data. Elasticities based on the total amount of work performed are close to twice the size of elasticities based on worker-level estimations. What is behind this difference is an essential topic for further investigation.

As pointed out by Barzel (1973) and Fehr and Goette (2007) time spent at work and effort are not necessarily the same. When interpreting our estimates, it is essential to remember that we measure effort, that is the number of HITs completed, rather than time spent, and employers are presumably interested in the elasticity of effort.[15] We can clearly rule out the negative elasticity of effort found in some prior research (Camerer, Babcock, Loewenstein, and Thaler, 1997; Fehr and Goette, 2007).

Finally, two important questions for future research arise from our result. First, why do the elasticities differ between the two jobs? The prior research has ascribed the variation in elasticities across studies to differences in the functioning of the markets examined, but our two jobs were in the same labor market and by the same employer. Hence, the differences in elasticities in this paper originate in the jobs and workers' response to them, which suggests that there is a vital role for worker attributes such as skills and preferences

[15] We cannot directly measure the quality of the responses, but if the quality is increasing in pay as the efficiency wage theory suggests, our estimates are lower-bound estimates of the "quality-adjusted" elasticities facing the employer.

in explaining differences in elasticities. Currently, we know little about how the distribution of skills and preferences affect labor supply elasticities, and this is an important question for future research. One promising way to frame this research is the monopsonistic competition labor market model, where employers offer jobs with different non-wage job characteristics and workers have different preferences for job attributes (Bhaskar and To, 1999; Manning, 2011).

Second, why are the elasticities not infinite? Mechanical Turk has many of the characteristics we associate with perfectly competitive labor market: many employers and workers, little or no discrimination, complete freedom to work as much or as little on each job as one wants, and no regulation. There are, however, two significant differences from our standard model of perfectly competitive markets. First, the offered jobs differ in content and design, and workers vary in preferences and skills, and therefore opportunity cost, as we just discussed. Second, workers have imperfect information about the offered jobs and whether the return to working on a job is sufficiently high, which imply that, even if the cost of entry and exit to a job is low, search costs are potentially still significant. Since these two explanations have different welfare implications, understanding the relative role of search costs and monopsonistic competition in determining the elasticity that employers face is an important topic for future research and essential for predicting the effects of policies such as minimum wage legislation with the spread of the "gig economy."

References

Barzel, Y. (1973): "The Determination of Daily Hours and Wages," *The Quarterly Journal of Economics*, 87(2), 220–238.

Bhaskar, V., and T. To (1999): "Minimum Wages for Ronald McDonald Monopsonies: a Theory of Monopsonistic Competition," *The Economic Journal*, 109(455), 190–203.

Blundell, R., and T. MaCurdy (1999): "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3, Part A, pp. 1559–1695. Elsevier, Amsterdam.

Blundell, R., T. MaCurdy, and C. Meghir (2007): "Labor Supply Models: Unobserved Heterogeneity, Nonparticipation and Dynamics," vol. 6, Part A of *Handbook of Econometrics*, chap. 69, pp. 4667 – 4775. Elsevier.

Buhrmester, M., T. Kwang, and S. Gosling (2011): "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?," *Perspectives on Psychological Science*, 6, 3–5.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *The Quarterly Journal of Economics*, 112(2), 407–441.

Chetty, R. (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3), 969–1018.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011): "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *The Quarterly Journal of Economics*, 126(2), 749–804.

Chetty, R., A. Guren, D. S. Manoli, and A. Weber (2011): "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," Working Paper 16729, National Bureau of Economic Research.

Dal Bó, E., F. Finan, and M. A. Rossi (2013): "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 128(3), 1169–1218.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (Forthcoming): "Monopsony in Online Labor Markets," *American Economic Review: Insights*.

Falch, T. (2010): "The Elasticity of Labor Supply at the Establishment Level," *Journal of Labor Economics*, 28(2), 237–266.

——— (2011): "Teacher Mobility Responses to Wage Changes: Evidence from a Quasi-natural Experiment," *American Economic Review*, 101(3), 460–65.

Fehr, E., and L. Goette (2007): "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97(1), 298–317.

French, S., and T. Stafford (2017): "Returns to Experience and the Elasticity of Labor Supply," Discussion Papers 2017-15, School of Economics, The University of New South Wales.

Goldberg, J. (2016): "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi," *American Economic Journal: Applied Economics*, 8(1), 129–49.

Heckman, J. J. (1993): "What Has Been Learned About Labor Supply in the Past Twenty Years?," *The American Economic Review*, 83(2), 116–121.

Kaplan, E. L., and P. Meier (1958): "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53(282), 457–481.

Keane, M., and R. Rogerson (2015): "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective," *Annual Review of Economics*, 7(1), 89–117.

Keane, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

Manning, A. (2011): "Imperfect Competition in the Labor Market," vol. 4 of *Handbook of Labor Economics*, chap. 11, pp. 973 – 1041. Elsevier.

Matsudaira, J. D. (2014): "Monopsony in the Low-Wage Labor Market? Evidence from Minimum Nurse Staffing Regulations," *The Review of Economics and Statistics*, 96(1), 92–102.

Oettinger, G. S. (1999): "An Empirical Analysis of the Daily Labor Supply of Stadium Venors," *Journal of Political Economy*, 107(2), 360–392.

Pörtner, C. C., N. Hassairi, and M. Toomim (2015): "Only if You Pay Me More: Field Experiments Support Compensating Wage Differentials Theory," Working paper, Seattle University.

Sokolova, A., and T. A. Sorensen (2018): "Monopsony in Labor Markets: A Meta-Analysis," IZA Discussion Paper 11966, IZA – Institute of Labor Economics, Bonn, Germany.

Staiger, D. O., J. Spetz, and C. S. Phibbs (2010): "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment," *Journal of Labor Economics*, 28(2), 211–236.

A Appendix

Figure A.1: Listing of jobs on Mechanical Turk