Labor Supply Elasticities in a Low Friction Labor Market*

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

Nail Hassairi
Department of Economics
University of Washington

October 2015

PRELIMINARY

*THE RESULTS PRESENTED ARE PRELIMINARY. PLEASE DO NOT QUOTE OR SHARE WITHOUT PERMISSION. This research was conducted under IRB approval from University of Washington and Seattle University. This 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.

Abstract

This paper estimates extensive and intensive margin labor supply elasticity using data from field experiments conducted in an online labor market. Contrary to prior analyses using micro data, we find that the intensive margin elasticities are more than twice the size of extensive margin elasticities and that both are substantial, even if conditioning on working. Furthermore, using data on all workers in the experiments whether they decide to work on our experiments or not, we find overall elasticities that range from 1.2 to 2.9, depending on experiment and specification. We argue that our results are consistent with the idea that off-line labor markets are characterized by frictions that lower elasticities and may reverse the ordering of extensive margin and intensive margin elasticities.

JEL codes: J2, J3, C93

1 Introduction

A persistent question in economics is the size of labor supply elasticities. Labor supply elasticities are of interest in their own right, but especially because they impact how we evaluate a variety of policies, first and foremost tax policy. There are two stylized "facts" that have fed the large literature on labor supply elasticities. First, estimates of labor supply elasticities using micro data are small, especially compared to what is implied by macro models (Blundell and MaCurdy, 1999; Chetty, Guren, Manoli, and Weber, 2011; Keane and Rogerson, 2015). Second, intensive margin elasticities are substantially lower than extensive margin elasticities (Heckman, 1993; Chetty, 2012). Neither of these conform nicely to economic models, which has led to a large literature trying to explain them away.[1]

Part of the explanation for the divergent results may be that using observed wages to estimate labor supply elasticities is associated with a number of potential data and econometric issues that lead to biased estimates of the effects of wages on labor supply. First, individual wages may be endogenous if, for example, unobservable preferences for work are correlated with unobserved productivity and therefore wages (Blundell, MaCurdy, and Meghir, 2007). Second, wages are often measured with substantial error, especially if wage per hour is calculated from self-reported hours worked and pay. Third, wage change may come from shifts in either labor supply and labor demand or both, and the driver of these shifts may unobservable (Oettinger, 1999; Farber, 2014). Finally, wages are only observed for workers who decide to work at a given job and for a given wage.

We therefore take a different approach from the previous literature. Instead of trying to infer labor supply elasticities from observed wages, we run field experiments that allow us to directly estimate labor supply elasticities. We offer jobs, randomly allocating arriving workers to different wages within each job, and observe workers' decision on whether to work or not and amount of work supplied. There are two main contributions. First, we show how elasticities behave in an environment that is close to the standard neoclassical model. Second, we can estimate labor

[1] The literature is too large to survey here. For a recent review see Keane (2011).

supply elasticities while controlling for selection because we observe all workers looking at our jobs, whether they work or not. Our experimental setup allows us to estimate extensive margin elasticities and intensive margin elasticities without encountering the econometric problems normally associated with estimating elasticities from regular labor market data. We show that labor supply elasticities are substantial in this market and that intensive margin elasticities are double that of extensive margin elasticities.

This 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 jobs offered, including pay. We offer two separate jobs on Mechanical Turk: One asks workers to tag images with keywords and the other asks them to write letters. Each job requires different skill sets and appeal to workers with different interests, thereby providing more general validity to our experiments. As described in Pörtner, Hassairi, and Toomim (2015), in each experiment we randomize the job characteristics that workers are presented with and the pay offered. Here we use the randomized pay to we estimate differences in work supplied for different levels of pay.

Mechanical Turk has three major advantages when estimating labor supply elasticities. First and foremost, we can experimentally vary the wage. This means that offered wages are exogenous to worker characteristics and we do not have to worry about supply or demand shifts affecting wages. Especially, workers entering or leaving our job in no way impact anybody else's wage. Second, there is substantially less scope for measurement error than in prior studies. We set the wage and we directly observe whether a worker accepts or rejects a job offer, and the amount of work supplied if the job offer is accepted. Finally, we observe offered wage independently of whether a worker decides to work or not.

In addition to data and econometric issues, a number of reasons have been suggested for why estimated labor supply elasticities are small and for why intensive margin elasticities are larger than extensive margin elasticities. The three are investments in human capital, credit constraints, and frictions in labor markets (Imai and Keane, 2004; Domeij and Flodén, 2006; Chetty, 2012). Mechanical Turk is especially of interest because it is less rigid than most off-line labor markets and therefore allow us to understand how labor supply elasticities perform in a low friction environment. There are low fixed and search costs on both sides and only short-term jobs are offered. Workers who find a job unattractive simply move on to another job without occurring any penalties, and employers engage in either minimal or no initial screening of workers. We can also rule out human capital accumulation as a factor that may impact our estimated elasticities.

Mechanical Turk does have a couple of downsides when estimating labor supply elasticities. First, we have little to no information about individual workers. We do present results based on survey information about basic demographic characteristics, but opposite our overall results who responds depends potentially on unobserved characteristics. Second, we cannot observe overall behavior for workers on Mechanical Turk, only how they perform in our experiment. We return to this later.

This work is closely related to Oettinger (1999) work on labor supply elasticities for stadium workers, the work by Camerer, Babcock, Loewenstein, and Thaler (1997) and Farber (2014) on taxi drivers in New York City, Barmby and Dolton (2009) on archeological dig, and Goldberg (Forthcoming) work on rural labor supply in Malawi.

We find strong positive elasticities, so little evidence of the negative wage elasticities found by Camerer, Babcock, Loewenstein, and Thaler (1997) or the negative elasticity of effort in Fehr and Goette (2007).

Our estimates of labor supply elasticities are decidedly short-run. The longest job ran only over 6 days.

It is possible for us to show large effects in what is effectively a standard neoclassical labor market.

2 Experimental Design

Amazon's Mechanical Turk is the largest of the emerging micro-task markets with over 100,000 registered workers from over 100 countries (Buhrmester, Kwang, and Gosling, 2011). Workers have to be 18 years or older, but otherwise there are few restrictions on participation. Work is paid per task rather than per hour—the corresponding hourly wage is lower than the overall US labor market, but will be close to the U.S. minimum wage. Individual tasks in a job are called HITs (Human Intelligence Tasks) and workers choose jobs from a list on the website that can be sorted by criteria such as pay per HIT and posting date.[2] Workers can preview a job before accepting it, and abort the job without penalty at any time. Between 5,000 and 30,000 HITs are completed each day (Ipeirotis, 2010). The Mechanical Turk labor market is built to be low friction for workers, allowing them to quickly move between jobs and work as much or as little as they desire on a given job. That does not mean that it is costless to move between jobs; there are clearly still search cost, and for some jobs it is more difficult to assess the work burden per HIT up front than others. We will return to this below.

Anyone can register to post jobs on Mechanical Turk. Examples of jobs include transcribing audio recordings into text, reviewing products, rewriting paragraphs, labeling images, searching for information, data entry, and answering surveys. Mechanical Turk allows requestors to require skills and "certifications" of workers. Our only requirement is that the computer accessing our jobs must be in the U.S. This allows us to estimate consistent wage responses, while achieving a sufficient sample size. U.S. Mechanical Turk workers are similar to the U.S. Internet population, and the income distribution closely follows the distribution for the overall U.S. population (Ipeirotis, 2008). It is possible to circumvent our location restriction through the use of proxy servers, but Amazon requires that workers provide a US tax ID number if they use a computer that appears to be in the US, which significantly limits the usefulness of using a proxy server to access Mechanical Turk. Requestors can reject HITs for subpar work. Having HITs rejected neg-

[2] The tagline for Amazon's Mechanical Turk is "Artificial Artificial Intelligence" to emphasize that these are jobs that are done by people. Appendix Figure A.4 shows an example of a job listing on Mechanical Turk.

atively affect workers because requesters can exclude workers based on rejection rates (Horton, 2011).

We offered two separate jobs, which ran at different times during 2013 and 2014. The two jobs were designed to be attractive to different segments within the Mechanical Turk worker community and to require different skill sets. In one job we offered workers a pictures tagging task, and in the other asked them to write short letters. This allows us to compare how labor supply elasticities respond to different types of work and thereby achieve broader validity. The experiments were originally designed to test the compensating wage differentials theory by randomizing offered combinations of job characteristics and pay (Pörtner, Hassairi, and Toomim, 2015). Here we use the randomization of pay offered.

Data collection begins as soon as a worker clicks on our offered job in the job listing. To ensure that workers who show up at different times of the day are equally likely to be presented with all job characteristics, we listed all the possible combinations in random order. Each worker that looks at our job is automatically assigned the next combination in the list. We observe whether the worker accepts the job and, if so, how many HITs are performed. Workers were not informed that the offered jobs were part of an experiment and were always presented with the same set of circumstances based on their unique worker ID number assigned by Mechanical Turk. 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 their output is potentially being monitored, but this monitoring is identical across the experiments and akin to what one would find in any job. The experiments were conducted exclusively through computers ruling out any experimenter bias.

Requestors can only contact workers they have paid in the past. We therefore paid all new workers a $0.25 "bonus". We do this only the first time a worker looks at one of our jobs; otherwise the worker is taken straight to the regular job. The bonus allows us to contact workers for our survey independently of whether they completed any real HITs or not. The bonus may make workers feel an obligation to work, which would inflate the number who do at least one

HIT and the number of HITs performed. This may bias upward our estimates of extensive and intensive margin elasticities. [TK run estimations for those who did not get a new worker bonus and those who did]

The image tagging job is straightforward and similar to many other tagging jobs offered on Mechanical Turk, where requestors have worker go through images before deciding which ones to license. Once a worker accepts 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.

Figure 1: Image Tagging Experiment

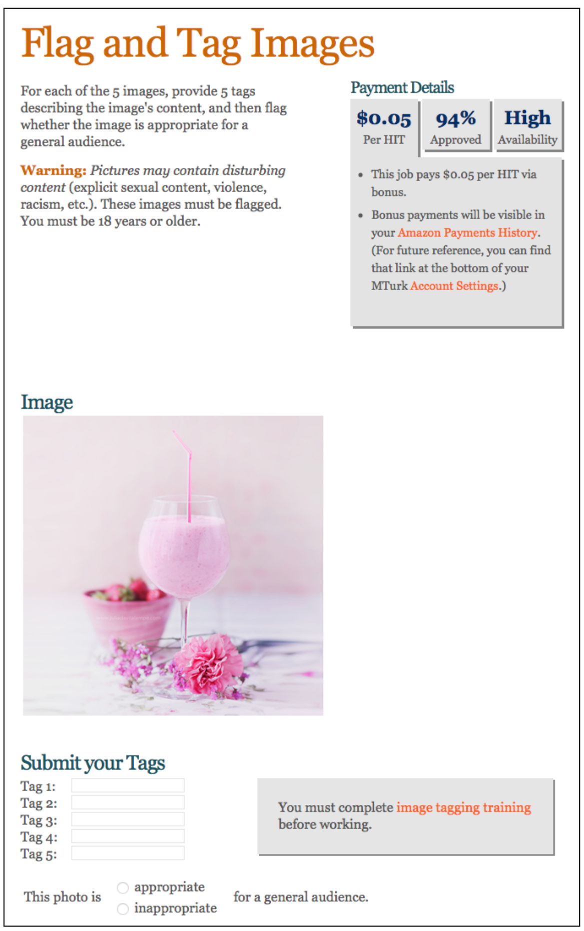
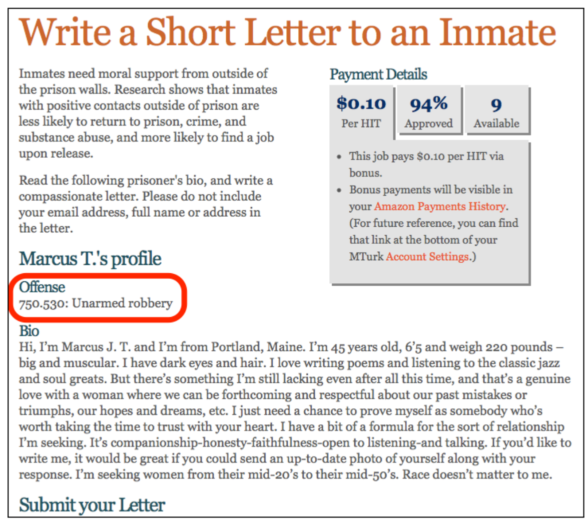


Figure 2: Letter Writing Experiment



Workers were randomly assigned to a pay per five images tagged, equal to 25 tags, of between USD 0.05 and USD 0.50 in USD 0.05 increments. The experiment ran over a six day period in 24 hour segments from 07.58 GMT. A worker would see one set of conditions during each 24 hour period and then after 07.58 GMT the worker job conditions and pay would be randomized anew. The randomization on subsequent days does not take account of previous job characteristics or pay that the worker has experienced. We choose 07.58 GMT because that was the time of the day where there were the fewest number of workers on Mechanical Turk.[3] This set-up allows us to both look at initial choice about labor supply and what determines the decisions to return on subsequent days and amount of work provided.

In the letter writing job the basic task is to write a positive and supportive letter to a prison inmate. An example is shown in Figure 2. The pay vary in 10 cent increments from USD 0.1 to USD 1.0 per HIT completed. As for the image tagging job this pay remained with the worker throughout the experiment. The letter experiment ran only through one 24 hour segment.

3 Estimation Strategy

Our experimental setup allows us to examine how wage affects the amount of work, H, supplied. Because we observe all workers, whether they reject or accept our jobs, we can directly model the selection into work and amount of work supplied. We first estimate the effect of offered wage and job characteristics on the decision to work:

$$1[H_i > 0] = \alpha + \beta_1 \log(w_i) + \mathbf{c}_i \beta_2 + \epsilon_i, \tag{1}$$

where $1[H_i > 0]$ is an indicator variable that takes the value 1 if the worker complete at least one HIT and 0 otherwise, w_i is observed pay per HIT for worker i, and $\mathbf{c_i}$ is a vector of job characteristics.[4] We estimate this extensive margin decision using both a linear probability model and a Logit model.

[3] A worker working around the change point could potentially see one set of condition initially and another set later on.

[4] We do not show the effects of job characteristics here. See Pörtner, Hassairi, and Toomim (2015) for those results.

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

$$H_i = \alpha + \beta_1 \log(w_i) + \mathbf{c}_i\beta_2 + \epsilon_i \text{ if } H_i > 0. \quad (2)$$

We present both OLS results and results from a censored model that takes into account upper bound censoring.

Finally, we estimate a censored regression model that takes into account both the lower bound censoring that occur at zero HITs when workers reject our job and the upper bound censoring built into the experiment.[5] The censored regression model implicitly requires two assumptions: that wages are observed for all workers independent of whether they work or not, and that wages are exogenous to the workers' labor supply. Neither assumption would be acceptable in standard labor market data, but are appropriate here. The experimental design provides an offered wage for all workers, whether they work or not, and this wage is by design exogenous to the labor supply because of randomization. The censored regression model also implies an assumption of no fixed cost of participation. In our case there are no fixed costs of work, or rather, they have already been incurred by the worker by joining Mechanical Turk (buying computer and internet connection and signing up for Mechanical Turk) and there are no fixed costs specific to our job.[6]

We calculate three wage elasticities for each experiment: the extensive margin elasticity, the intensive margin elasticity conditional on working, and the "overall" elasticity for all workers whether they work or not. One question is the extent to which workers consider the offered wage a transitory wage change, in which case the elasticities estimates are Frisch elasticities. If workers see the offered wage as transitory we would expect no substantial income effect. Our expectation is that workers on Mechanical Turk treat each offered job as temporary and that

[5] In the cases where there are only one lower and one upper bound censoring point, the results from this estimation will be the same as that from a Tobit model.

[6] For a more detailed discussion of the three assumptions see Blundell, MaCurdy, and Meghir (2007).

if a job shows up which pays more than expected they treat this as a temporary shock since most active jobs on Mechanical Turk are relatively short-lived. Furthermore, all other jobs on Mechanical Turk remains at the same wage.

The extensive margin elasticity captures the effect of wage changes on the probability of working on our given job. The coefficient on the extensive margin regressions do not directly show the extensive margin elasticity, but we can calculate it as

$$\epsilon_e = \frac{\partial \Pr[H > 0] / \Pr[H > 0]}{\partial w / w} = \frac{\beta_1}{\Pr[H > 0]}, \tag{3}$$

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. This requires picking a number for the probability; here we use the participation rate for each experiment.[7]

Similarly, we calculate the intensive margin elasticities as

$$\epsilon_i = \frac{\partial H / H}{\partial w / w} = \frac{\beta_1}{H}, \tag{4}$$

using the results from equation (2) with log wage as the explanatory variable, where H is the number of HITs completed. For the intensive margin conditional on working we use the number of HITs completed by those who did at least one HIT for the experiment, whereas for the intensive margin for all workers we use the average over all workers.[8]

Fixed effects estimations has been used to overcome problems that arise from unobservable worker characteristics (see, for example, MaCurdy, 1981; Browning and Angus Deaton, 1985; Altonji, 1986; Oettinger, 1999). The idea is that observing the same worker, in principle, allows us to eliminate unobservable worker traits that drive selection into jobs. There are three drawbacks to this approach. First, it requires workers that receive different wages. Second, if those who

[7] If wage enter linearly in labor supply function the formula is $\beta_1 \frac{w}{\Pr[H>0]}$.

[8] Dividing by the average number of HITs for the censored regression ignores that the latent number of HITs supplied is lower than the observed when setting those who work to zero.

change wages is a non-random sample of workers there can still be important selection effects that will bias the results. In our case, although fixed effects will remove time-invariant worker characteristics, we may still have workers who self-select out of work because their reservation wage is above the offered wage. These workers will remain unobserved, potentially leading to biased results even with longitudinal data. We varied our offered wage substantially more than it would be possible to do in a regular off-line labor market to lower the chance of self-selection, but workers may still decide not to work because we are not offering them enough to surpass their reservation wage. Finally, fixed effects exacerbate any measurement errors in the data.

We ran the image tagging experiment over six days to examine whether the fixed effects approach provides comparable results to our experimental results. Workers were presented with a randomly allocated set of conditions and pay each day they visited the job. Although the conditions and wage that a worker face are randomized anew each day, only the conditions on the first day a worker visits can truly be considered random. Conditions may affect a worker's likelihood of looking at our offered job again. Hence, all data collected after the first day a worker visits are potentially affected by self-selection.

To compare with the first visit results, we estimate how wage affects amount of work done using fixed effects. We first estimate the extensive margin

$$1[H_{it} > 0] = \alpha + \beta_1 \log(w_{it}) + \mathbf{c}_{it}\beta_2 + \mu_i + \epsilon_{it}, \tag{5}$$

where i is the individual worker, t is visit number, and μ_i is a time invariant worker fixed effect, for all days where the worker looked at our job using a regular linear fixed effects model and a conditional Logit model. We then estimate the intensive margin model using only those workers who complete at least one HIT on their visit

$$H_{it} = \alpha + \beta_1 \log(w_{it}) + \mathbf{c}_{it}\beta_2 + \mu_i + \epsilon_{it} \text{ if } Y_{it} > 0. \tag{6}$$

Finally, we estimate the same model but include all worker-visit observations, including those where the worker did no HITs. Neither of these two models take into account the censoring at zero HITs and the upper level censoring in the experiment.[9]

Following the experiments we surveyed workers who visited our image tagging experiments. To examine whether worker characteristics affect the results, we estimate the first day models—effect of wage on effort supplied—using the survey responses. Obviously whether somebody responds to the survey is not a random event. Hence, the results using the survey respondent can best be thought of an additional way to examine how selection can impact the results when examining elasticities.

4 Results

Job characteristics and pay are only truly random the first time a worker visits the job. For the image tagging experiment we therefore initially focus only on the first time a worker was observed, and return to what can be learned from the panel aspect below.[10] During the image tagging experiment's six 24 hour segments, 4,311 workers visited the job.[11] The letter writing experiment ran for one 24 hour segment and 2,111 workers visited. Figure 3 shows the distribution of work done in each experiments. The panels on the left include those who chose not to work, while the panels on the right are conditional on working to show the distribution of work done more clearly.

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. For the letter writing experiment 73 percent did not work,

[9] There are methods that allow for fixed effects in censored regression models, but the purpose of this paper is to evaluate the standard models used to examine the compensating wage differentials theory, rather than evaluate the different methods available. See Dustmann and Rochina-Barrachina (2007) for a comparison of different selection correction models for panel models.

[10] The first day is not necessarily the first day the experiment ran, but rather the first day we observed the worker in the image tagging experiment.

[11] We tried to run the image tagging experiment about seven months prior, but aborted it within hours because of server load issues. Removing workers who showed up for both has no effect on our results. The long period between the aborted attempt and the final run was partly because of the time required to design and run load testing programs for the servers and partly to minimize contamination between the aborted run and the final experiment.

Letter writing experiment

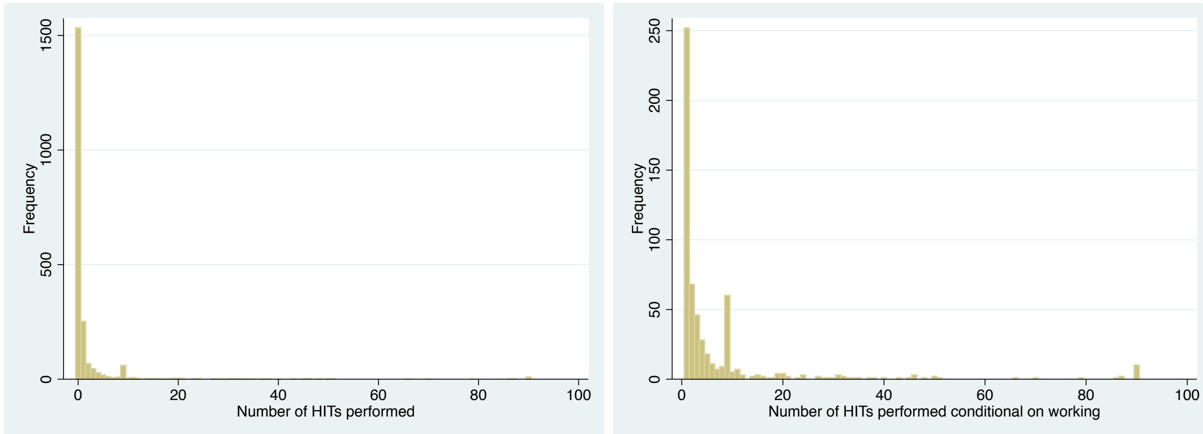


Image tagging experiment

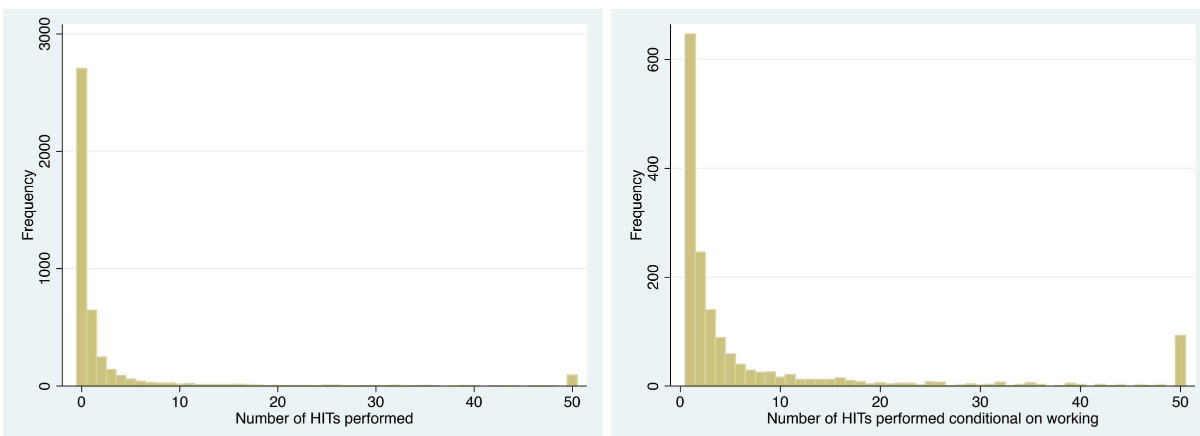


Figure 3: Distribution of work done by experiment

leaving 578 workers who completed one or more HITs The advantage of Mechanical Turk is that we observe all the workers who decide that they do not want to work, and who would not be observed in standard labor markets.

One of the job characteristics tested in the experiment was availability and for the letter writing experiment workers in the low availability condition were not allowed to work more than 9 HITs, whereas all others have an upper limit of 90 HITs. Both show clearly in the histogram. In the image tagging experiment the low availability was not implemented as a fixed cut-off, so the only limit is the maximum of 50 HITs. Almost 100 workers reached the maximum on their first day working on the image tagging experiment.

In total, 4,366 letters were written and 60,695 images tagged—equal to 303,475 keywords on the first day. The pay-outs to workers were \$3,055.1 and \$3,808.3.

Tables 1 show estimated effects of wage on extensive and intensive margins for the two experiments. In each, the first two columns are on extensive margin using a linear probability model and a logit model, the third and fourth columns are on intensive margin using the sample of workers who completed at least one HIT, and the final column show results when taking account of censoring at zero and above using all workers. Each table shows results using wage and log wage separately.

Table 2 shows the elasticities from the different models. One of the characteristics of the Mechanical Turk labor market is the low adjustment cost.[12] If a worker dislike the offered set of wages and job characteristics it is easy to move on to another job. Workers do, however, still have imperfect information about the job. They may, for example, not know how long it takes to complete a HIT, and therefore not know whether it is worthwhile to work on it. This suggests that many workers will complete one HIT and then decide whether they want to continue working. We therefore also show estimated elasticities using one HIT as the cut-off point. Extensive margin results are for the probability of working more than one HIT, and intensive margin results are the number of HIT completed minus one.

The extensive margin wage elasticities are 0.13 for the image tagging experiment and 0.27 for the letter writing experiment. For comparison, workers on an archeological dig in Syria in the 1930s were found to have an extensive margin elasticity of 0.035, a labor market experiment in Malawi showed an elasticity of around 0.15, whereas stadium vendors in the US had an elasticity of between 0.55 and 0.65 (Barmby and Dolton, 2009; Goldberg, Forthcoming; Oettinger, 1999).

Although our extensive margin elasticities are in the range of published results they are lower than other U.S. extensive margin elasticities; both Oettinger (1999) and Chetty (2012) find elasticities that are more than twice as large as ours. Three features of the Mechanical Turk labor market combine to explain our lower estimates. First, workers do not commit to a particular

[12] See Chetty (2012) for a discussion of how to estimate elasticities in the presence of adjustment costs and frictions.

Table 1: Effects of Wage on Extensive and Intensive Margins

	Extensive		Intensive	
	Worked = 1, not = 0		Number of HITs Performed	
	LPM	Logit	Censored	
Sample	Full[a]	Full[a]	Worked[c,e]	Full[d,f]
Image Tagging Experiment—Zero HIT Cut-off				
Log wage	0.048***	0.220***	2.820***	3.277***
	(0.011)	(0.049)	(0.514)	(0.484)
Observations	4,311	4,311	1,605	4,311
Mean of dependent variable	0.372	0.372	7.6	2.8
Image Tagging Experiment—One HIT Cut-off				
Log wage	0.050***	0.315***	3.877***	5.713***
	(0.009)	(0.058)	(0.841)	(0.852)
Mean of dependent variable	0.222	0.222	11.996	2.816
Observations	4,311	4,311	958	4,311
Letter Writing Experiment—Zero HIT Cut-off				
Log wage	0.073***	0.400***	4.962***	6.175***
	(0.014)	(0.076)	(1.074)	(0.934)
Observations	2,111	2,111	578	2,111
Mean of dependent variable	0.274	0.274	7.6	2.1
Letter Writing Experiment—One HIT Cut-off				
Log wage	0.057***	0.496***	8.198***	11.062***
	(0.011)	(0.099)	(1.907)	(1.859)
Observations	2,111	2,111	326	2,111
Mean of dependent variable	0.154	0.154	12.620	2.068

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

[a] Sample consists of all workers on the first day they are observed during the experiment, whether they worked or not.

[c] Sample consists of workers who worked, i.e. completed at least one HIT, on the first day the worker was observed during the experiment. Of the 1,605 workers who worked on the first day they were observed, 92 were right-censored observations.

[d] Of the 4,311 observations, 2,706 were left-censored observations, 1,513 uncensored observations, and 92 right-censored observations.

[e] Sample consists of workers who worked, i.e. completed at least one HIT, on the first day the worker was observed during the experiment. Of the 578 workers, 68 were right-censored.

[f] Of the 2,111 observations, 1,533 were left-censored observations, 510 uncensored observations, and 68 right-censored observations.

Table 2: Wage Elasticities of Labor Supply

	Image tagging		Letter writing	
	0 HIT[a]	1 HIT[b]	0 HIT[a]	1 HIT[b]
Extensive margin	0.13	0.23	0.27	0.37
Intensive margin, conditional on working[c]	0.37	0.32	0.65	0.65
Intensive margin, all workers	1.17	2.03	2.94	5.35

Note. Each elasticity calculated at the mean value of the dependent variable.

[a] Based on results from Table 1 with zero HITs as cut-off.

[b] Based on results from Table 1 with one HITs as cut-off.

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

number of HITs by working one HIT. Second, fixed entry costs for individual jobs are low once a worker is on the Mechanical Turk platform. Finally, workers likely have imperfect information about how attractive an offered combination of pay and job characteristics is, leading many workers to try the job as part of their search for the jobs with sufficiently high return to effort. If workers routinely do a "trial" HIT before deciding whether to continue working we should see much higher elasticities if we ignore the first HIT. The elasticities using one HIT as the cut-off are, indeed, 10 percentage points higher than the pure extensive margin elasticities.

The commonly held view is that intensive margin elasticities are substantially lower than extensive margin elasticities (Heckman, 1993). For the sample of workers who worked intensive margin elasticities are, however, more than twice as large as our extensive margin elasticities—using zero HITs as the cut-off. The intensive margin elasticities are 0.37 for the image tagging experiment and 0.65 for the letter writing experiment.

The standard available labor market data only allows for the calculation of the intensive margin elasticities for those workers who are working. We, however, have offered wages for all workers independently of whether they decide to work or not. We can therefore also calculate intensive margin elasticities for the sample of all workers who looked at our offered jobs. These intensive margin elasticities are large. The lowest is for the image tagging experiment with zero HITs as the cut-off and even that is 1.2. The equivalent one for the letter writing experiment is almost 3, and the largest is for the one HIT cut-off for the same experiment at 5.4.

Chetty (2012, p 1009) argues that the commonly held view that the extensive margin elasticities are larger than intensive margin elasticities may simply be because of frictions and that "[i]n steady state, the intensive elasticity may actually be larger than extensive elasticities…" That our intensive margin elasticities are larger than our extensive margin elasticities supports his argument and, indirectly, our conjecture that the Mechanical Turk labor market is a low friction labor market. Although our elasticities are not directly comparable to other published numbers given the low entry and exit costs of the individual Mechanical Turk jobs and that a worker does not commit to a particular number of HITs when entering, our results do suggest that workers on Mechanical Turk behave as we would expect given standard labor economics models and that the Mechanical Turk labor market is close to the standard neo-classical model of labor supply.

4.1 Worker Fixed Effects Results

Table 3 shows fixed effects estimates for both extensive and intensive margin. The selection over time shows up in the larger mean of the higher percentages of people who work compared to the first day analysis. In the panel data, 42 percent of workers work—up from 37 for the first day analysis—and the average number of HITs performed per worker per day is almost twice as large as in the single day data.

Table 3: Effects of Wages on Extensive and Intensive Margins—Worker Fixed Effects

	Extensive		Intensive	
	Worked = 1, not = 0		Number of HITs Performed	
	Linear	Logit	Linear	linear
Sample	Full[a]	Full[b]	Worked[c]	Full
Log wage	0.113***	0.885***	7.365***	4.562***
	(0.009)	(0.076)	(0.552)	(0.259)
Observations	7,954	2,357	3,330	7,954
Number of workers	4,311	719	1,830	4,311
Mean of dependent variable	0.419	0.542	13.095	5.482
Elasticities at mean values	0.270		0.562	0.832

Note. Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.
[c] This sample has a higher number of people than than the first day results because there are 125 workers that did not work on the first day they visited the job, but did work on a subsequent day. Hence, the first day number of observations for the intensive margin is 1605, whereas it is 1830 for the fixed effects estimations on intensive margin.

Compared to the first day results, the extensive margin results show substantially stronger effects of wage on the likelihood of working. In the linear model, the effect of wage is twice as large in the panel data than in the first day data. For the intensive margin only the linear results are directly comparable with the first day results. Again we see a substantially larger effect of wage, which the estimated effect on number of HITs performed, conditional on working, are 2 to 3 times larger for the fixed effects results than for the first day results.

The fixed effects extensive margin elasticity are just over twice as large as using only the first day data, which puts it close to the one HIT cut-off results. Similarly, the two intensive margin elasticities are also larger than in the first day results. A potential reason they increase less than for the extensive margin may be that the fixed effects estimations do not control for censoring they way we did for the frist day results. That said, conditional on working the intensive margin elasticities is 0.562 and using all workers it is 0.832.

4.2 Worker Characteristics and Their Effects

A downside of Mechanical Turk is the lack of background information on workers, including how long their tenure on Mechanical Turk is. We can, however, create measures for how long workers have been on Mechanical Turk, based on prior experiments and the experiments here. Experience on Mechanical Turk is especially of interest since Farber (2014) argues that taxi drivers in New York City are more likely to be optimizers the longer they have been working. Hence, it may be that workers who have been on Mechanical Turk longer may be more "rational" and therefore less likely to try the HIT unless they find it worthwhile.

Tables 4 and 5 show extensive and intensive margins results, controlling for whether we have observed a worker before and when. Our earliest experiments on Mechanical Turk ran in September 2010 and January 2011 (Toomim, Kriplean, Pörtner, and Landay, 2011). Of the workers in those experiments, only 45 show up among workers who looked at our letter writing experiment in March 2014, and only 55 show up at our image tagging experiment in November 2014. We therefore combine these workers with workers we first observed at another experiment that ran in June 2013 to form the dummy variable "First observed June 2013 or before", which has a total of 205 workers in the letter writing experiment and 235 workers in the image tagging experiment. Finally, the letter writing experiment ran in March 2014 and in April 2014 we had an initial run of the image tagging experiment that was aborted within hours of starting because of server load issues. A total of 439 workers in the image tagging experiment where first observed in one of these two experiments. Hence, only about 10% of the workers in the letter writing experiment and 15% of the workers in the image tagging experiment were workers we had seen before.

Table 4: Effects of Job Characteristics on Extensive and Intensive Margins for Letter Writing Experiment Controlling for Experience on Mechanical Turk

	Without Interactions			With Interactions		
	Extensive Worked = 1	Intensive HITs Performed		Extensive Worked = 1	Intensive HITs Performed	
	LPM	Censored		LPM	Censored	
Sample	Full[a]	Worked[b]	Full[c]	Full[a]	Worked[b]	Full[c]
First observed June 2013 or before	−0.007	2.409	0.175	0.044	−1.448	1.629
	(0.032)	(2.339)	(2.058)	(0.080)	(5.881)	(5.020)
Log wage	0.073***	4.960***	6.173***	0.075***	5.364***	6.462***
	(0.014)	(1.073)	(0.935)	(0.014)	(1.122)	(0.979)
Log wage × observed 2013				−0.006	−4.801	−2.497
				(0.049)	(3.716)	(3.257)
Observations	2,111	578	2,111	2,111	578	2,111
Dependent variable mean	0.274	7.6	2.1	0.274	7.6	2.1
Elasticities at mean values	0.266	0.653	2.940	0.274	0.706	3.077

Notes. Standard errors in parentheses; * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers on the first day they are observed during the experiment, whether they worked or not.
[b] Sample consists of workers who worked, i.e. completed at least one HIT, on the first day the worker was observed during the experiment. Of the 578 workers, 68 were right-censored.
[c] Of the 2,111 observations, 1,533 were left-censored observations, 510 uncensored observations, and 68 right-censored observations.

Controlling for experience on Mechanical Turk does little to change the estimated elasticities. Hence, less experienced workers behave as expected and show positive and substantial labor supply elasticities. If anything it appears that less experienced workers show a stronger response to wage changes than workers we have observed before.

5 Conclusion

Mechanical Turk is clearly not like “off-line” labor markets. There are no explicit contract, no set working hours, no commuting, and clothing is entirely optional. Is it, however, similar to the market for freelance or independent contractor work. Freelancing, independent contracting, and consulting work is rapidly becoming more and more important in the US economy. A recent estimate is that there are 17.7 million independent workers, making close to USD 1.2 trillion in

Table 5: Effects of Job Characteristics on Extensive and Intensive Margins for Image Tagging Experiment Controlling for Experience on Mechanical Turk

	Without Interactions			With Interactions		
	Extensive Worked = 1	Intensive HITs Performed		Extensive Worked = 1	Intensive HITs Performed	
	LPM	Censored		LPM	Censored	
Sample	Full[a]	Worked[b]	Full[c]	Full[a]	Worked[b]	Full[c]
First observed June 2013 or before	−0.204***	−1.322	−9.166***	−0.185*	−8.887	−5.355
	(0.031)	(1.992)	(1.628)	(0.102)	(6.894)	(5.454)
First observed March/April 2014	−0.226***	2.390	−8.822***	−0.250***	−1.872	−6.107
	(0.024)	(1.535)	(1.220)	(0.075)	(4.594)	(3.822)
Log wage	0.046***	2.807***	3.247***	0.042***	2.858***	3.039***
	(0.011)	(0.515)	(0.485)	(0.011)	(0.532)	(0.511)
Log wage × observed 2013				0.038	−1.123	3.209
				(0.047)	(3.415)	(2.707)
Log wage × observed 2014				0.024	−2.146	1.736
				(0.036)	(2.655)	(1.986)
Observations	4,311	1,605	4,311	4,311	1,605	4,311
Dependent variable mean	0.372	7.6	2.8	0.372	7.6	2.8
Elasticities at mean values	0.124	0.369	1.160	0.113	0.376	1.085

Notes. Standard errors in parentheses; * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers on the first day they are observed during the experiment, whether they worked or not.
[b] Sample consists of workers who worked, i.e. completed at least one HIT, on the first day the worker was observed during the experiment. Of the 1,605 workers who worked on the first day they were observed, 92 were right-censored observations.
[c] Of the 4,311 observations, 2,706 were left-censored observations, 1,513 uncensored observations, and 92 right-censored observations.

total income in 2013, and these numbers have been increasing over time (MBO Partners, 2013).[13] Finally, Mechanical Turk attracts people actively looking for work, rather than being a sample of undergraduate students participating in a lab experiment.

We find that labor supply elasticities follow the pattern predicted by the standard neoclassical model. Intensive margin elasticities when conditioning on working are about twice as larger as extensive margin elasticities. Because we can observe all workers whether they decide to work or not, we can also estimate intensive margin elasticities for all workers. These elasticities are approximately 10 times what we find for extensive margin elasticities and substantially larger what the previous literature has found.

[13] There is, however, substantial uncertainty about these numbers since the Bureau of Labor Statistics does not directly count these people.

References

Altonji, J. G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 94(3), S176–S215.

Barmby, T., and P. Dolton (2009): "What Lies Beneath? Effort and Incentives on Archaeological digs in the 1930's.," Working paper, University of Aberdeen.

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.

Browning, M., and M. I. Angus Deaton (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53(3), 503–543.

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., 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," NBER Working Papers 16729, National Bureau of Economic Research, Inc.

Domeij, D., and M. Flodén (2006): "The labor-supply elasticity and borrowing constraints: Why estimates are biased," *Review of Economic Dynamics*, 9(2), 242 – 262.

Dustmann, C., and M. E. Rochina-Barrachina (2007): "Selection correction in panel data models: An application to the estimation of females' wage equations," *The Econometrics Journal*, 10(2), 263–293.

Farber, H. S. (2014): "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," Working Paper 20604, National Bureau of Economic Research.

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.

Goldberg, J. A. (Forthcoming): "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi," *American Economic Journal: Applied Economics.*

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

Horton, J. (2011): "The condition of the Turking class: are online employers fair and honest?," *Economic Letters.*

Imai, S., and M. P. Keane (2004): "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 45(2), 601–641.

Ipeirotis, P. (2008): "Mechanical Turk: The Demographics," `http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html`, Accessed: 9/18/2009.

——— (2010): "Demographics of Mechanical Turk," *New York University Working Paper.*

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.

MacUrdy, T. E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89(6), 1059–1085.

MBO Partners (2013): "The State of Independence in America - Third Annual Independent Workforce Report," Discussion paper.

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.

Toomim, M., T. Kriplean, C. C. Pörtner, and J. A. Landay (2011): "Utility of Human-Computer Interactions: Toward a Science of Preference Measurement," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, pp. 2275–2284. ACM.

A Appendix

Figure A.4: Listing of jobs on Mechanical Turk