Testing the Compensating Wage Theory Using Online Labor Market 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

Michael Toomim
Computer Science
University of Washington

Nail Hassairi
Department of Economics
University of Washington

December 2014

PRELIMINARY

Abstract

The theory of compensating differentials dates back to Adam Smith's Wealth of Nations. A basic problem in testing the compensating wage differentials theory is that workers self-select into jobs with a given set of characteristics or amenities. To the extent that this self-selection is driven, at least partly, by unobservable worker or job characteristics that also influence the wage, simply regressing job characteristics on observed wage will provide biased estimates of how the job characteristics affect wage. Empirical tests have therefore been unable to find support for the compensating wage differentials theory for characteristics other than risk of death. To overcome the self-selection and unobservable job characteristics problem, we test the compensating wage theory using online labor markets. We set up and run a series of experiments designed to test the different aspects of Adam Smith's original ideas for what determines differences in wages. Our results strongly supports the compensating wage differentials theory. Job characteristics have significant effects on the extensive margin, but little or no effect on the intensive margin, which shows that self-selection into jobs with specific characteristics is, indeed, the most likely driver of the lack of support for the compensating wage differentials theory in the prior literature.

JEL codes: J3, J2

1 Introduction

> The five following are the principal circumstances which, so far as I have been able to observe, make up for a small pecuniary gain in some employments, and counterbalance a great one in others: first, the agreeableness or disagreeableness of the employments themselves; secondly, the easiness and cheapness, or the difficulty and expence of learning them; thirdly, the constancy or inconstancy of employment in them; fourthly, the small or great trust which must be reposed in those who exercise them; and, fifthly, the probability or improbability of success in them.
>
> Adam Smith (1776, Book I, Chapter X, Part I)

The theory of equalizing differences, also known as the compensating wage differentials theory, is one of the oldest economic theories, tracing its roots to Adam Smith's Wealth of Nations. The basic idea is that observed wage differences equalize differences in the characteristics of both work and workers. In the most common form of the theory, the wage difference reflects the consumption value of job characteristics. A professors may, for example, earn less than she would in a non-academic job because she have more control over both tasks engaged in and timing of work.

Despite many attempts at testing the theory, the only job characteristic consistently shown to affect wages is risk of death (Rosen, 1986). The basic problem in testing the compensating wage differentials theory is that workers may self-select into jobs with a given set of characteristics or amenities. If self-selection is at least partly driven by unobservable worker or job characteristics that also affect wage, simply regressing job characteristics on observed wage will lead to biased estimates of the effect of job characteristics on wage (Brown, 1980).

To overcome the self-selection and unobservable job characteristics problem, we test the compensating wage theory using online labor markets. We set up and run a series of experiments designed to test the different aspects of Adam Smith's original ideas for what determines differences in wages as shown in the quote above.

This work is made possible by the emergence of online labor markets for micro-tasks. There is an increasing number of micro-task markets, but Amazon's Mechanical Turk (www.mturk.com) is the largest. Employers are allowed to post almost anything as a job and a variety of different jobs are posted on Mechanical Turk, such as transcribing audio recordings into text, reviewing products, rewriting paragraphs, labeling images, searching for information, data entry, and answering surveys. Work is paid per task rather than per hour. Most jobs pay under a dollar, and many as little as one cent, per task completed.

The major advantage of online labor markets is that we control the different aspects of the jobs offered, such as job type, difficulty, and pay. For each worker we observe how many tasks, if any, she performs. Furthermore, we observe detailed information about each task performed, such as output and start and end times. Hence, conditional on workers looking at the offered job, we can examine how labor supply respond to job characteristics and thereby test the compensating wage differentials theory.

Two additional advantages of using online labor markets are the low search cost and short-term nature of the jobs offered. The fixed costs on both sides of the labor market are minimal, which allows for more variability in the labor response. Workers who do not find a job attractive simply move on to another job without occurring any penalties and employers engage in either minimal or no initial screening of workers. The low entry and exit costs of individual jobs, together with the short time required for each tasks, means that we can estimate both extensive and intensive margins for labor supply.

We offer two separate jobs on Mechanical Turk: One asks workers to tag images with key-words and the other asks them to write letters. Each job require different skill sets and appeal to workers with different interests, thereby providing more general validity to our experiments. Although the results from the two jobs are not directly comparable, using two different jobs also allow us to begin to understand how compensating wage differentials respond to job types. Within each job we vary four job characteristics that broadly correspond to four of the "principal circumstances" set out by Adam Smith in the quote above: agreeableness of the task, cost

4

of learning, availability, and probability of success.[1] In each experiment, we randomize the job characteristics that workers are presented with and the pay offered. For example, for agreeableness we randomly assign workers who look at our offered work to either an "agreeable" version of the task or a more "disagreeable" version of the same tasks together with a random pay offered. Combining random pay and different characteristics, we can estimate difference in work for a given pay and pay required to compensate workers for less desirable characteristics.

[RESULTS OVERVIEW]

2   Literature Review

[TO BE ADDED]

3   The Mechanical Turk Labor Market

Amazon's Mechanical Turk is the largest and most flexible of the emerging micro-task markets. Anyone can register to post jobs on Mechanical Turk and the main restriction for people looking to work is age. The individual tasks in a job are called HITs (Human Intelligence Tasks).[2] The suppliers of labor are "workers" and the agents demanding labor are "requesters." Mechanical Turk has over 100,000 registered workers from over 100 countries (Buhrmester, Kwang, and Gosling, 2011).

Each job on Mechanical Turk has a title and description, and the worker can click to preview a job before accepting it, and abort the job without penalty at any time. Workers choose jobs from a list on the website that can be sorted by criteria such as price and posting date, or searched by keyword or employer name. Work is paid per task rather than per hour, but the

---

[1] Adam Smith's idea of the amount of trust required essentially corresponds to the current idea of efficiency wages in modern labor markets (Shapiro and Stiglitz, 1985). The set-up required to test this is substantially different from the other four circumstances and we therefore plan to do that as a separate paper.

[2] The tagline for Amazon's Mechanical Turk is "Artificial Artificial Intelligence" to emphasize that these are jobs that are done by people.

equivalent hourly rate is commonly less than the U.S. minimum wage.[3] There are generally between 5,000 to 30,000 tasks completed each day (Ipeirotis, 2010). Workers communicate on a 3rd-party web forum, share tips, and discuss jobs and employers they had good or bad experiences with (see, for example, www.turkernation.com). Requesters can reject HITs for subpar work. Having HITs rejected has negative consequences for workers because requesters can exclude workers with high rejection rates (Horton, 2011).

The demographics of workers have been studied by posting surveys to Mechanical Turk itself (Ipeirotis, 2008). This research finds that 46% of workers are in the United States, with 34% in India, and 19% in other countries. The demographics of Mechanical Turk workers are similar to the Internet population, although slightly more female, slightly younger, and individuals tend to be single and with smaller families. Many report having Master's and Ph.D. degrees, and the income distribution closely follows the distribution for the overall U.S. population.

Mechanical Turk is clearly not like other, "real", 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, although the scale of the jobs obviously vary. 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 total income in 2013 and these numbers have been increasing over time (MBO Partners, 2013).[4] Finally, Mechanical Turk attracts people actively looking for work, rather than being a sample of undergraduate students participating in a lab experiment.

---

[3] The tax implications of working on Mechanical Turk are unclear, but Amazon does collect tax identification number from workers from both US and other countries.

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

4 Theory

To fix ideas, assume that two different types of jobs exists, indexed by $D = 0, 1$; jobs of type 0 are attractive, while jobs of type 1 are not.[5] Each worker's preferences are defined over two types of consumption goods and are represented by the utility function $u = u(C, D)$, where C is the consumption good purchased in the market and D is the consumption indicator of a job. We assume that for a given level of C workers will not prefer $D = 1$ to $D = 0$, or in terms of utility, $u(C, 0) \geq u(C, 1)$.

The utility function allows an exact calculation of how much income or market consumption the worker must be compensated to undertake the less preferred job. To see this, let C_0 be the market consumption when $D = 0$. Define C^* as the consumption level required to achieve the same utility on a $D = 1$ type job as C_0 guarantees on the $D = 0$ job. Then C^* satisfies $u(C^*, 1) = u(C_0, 0)$. Because $D = 1$ is never preferred to $D = 0$ if the consumption level is the same, it follows that $C^* \geq C_0$. Now define the difference $Z = C^* - C_0$ as the compensating variation for $D = 1$ compared with $D = 0$. This is the additional compensation necessary to make the worker indifferent between the two types of jobs.

This basic formulation of the compensating wage theory suggests a direct way of testing the theory: Randomly allocate potential workers to combinations of types of jobs and offered pay and observe whether there are statistically significant difference in the probability that workers accept the job for the same pay.

5 Experimental Design

Testing the compensating wage differentials theory requires translating the conditions described by Adam Smith into experimental set-ups. We set up two separate jobs to test how compensating wage differentials responds to different types of work and to achieve broader external validity. The two jobs are designed to be attractive to different demographics within the Mechanical

---

[5] This follows Rosen (1986).

Turk worker community and to require different skill sets. In one job the workers are offered a pictures tagging task, and in the other they are asked to write short letters. The idea is that the letter writing job has a higher "barrier to entry" than the image tagging job and that the individual tasks take longer for the letter writing job than the image tagging job. This allows us to compare extensive and intensive margins by job type.

Within each job we have four conditions that broadly correspond to the "principal circumstances" set out by Adam Smith: agreeableness of the task, cost of learning, availability, and probability of success. In each experiment/job we randomize the job characteristics that workers are presented with and the pay offered. The main difficulty is how to change these circumstances without fundamentally changing the job itself. We first describe the general set-up for the jobs and then each specific job and the implementation of the condition in detail below.

Figure 1 shows an example of listing of available jobs on Mechanical Turk. We act as a regular employer on Mechanical Turk and workers see our jobs as one of many offered HITs. If a worker click on our offered job she is randomly assigned to a treatment and data collection begins. We then observe whether she accepts the job and, if so, how many tasks she performs. In addition, we measure start times and end times for individual task and overall start and end times.

Worker are unaware that the offered jobs are part of an experiment and is always presented with the same set of circumstances based on the unique worker ID number assigned by Mechanical Turk. The reason for not informing workers that they are part of an experiment is to rule out an observer effect, where workers changes behavior in response to being part of an experiment. Workers do, however, clearly know that their work output is being monitored as part of the job, but this monitoring is identical across the experiments and akin to what one would find in any job. Finally, the experiment was conducted exclusively through computer interface ruling out any experimenter bias.

All new workers receive a USD 0.25 "bonus." The main reason is that we want to survey workers once the experiments are over but Amazon only allows requestors to contact workers

Figure 1: Listing of jobs on Mechanical Turk

that have completed and been paid for at least one tasks even though it may only involve clicking

a button like here. Figure 2 shows the new worker bonus offer provided to workers that we have

not previously paid. This is shown the first time a worker accepts one of our HITs; otherwise

the worker is taken straight to the regular job. A potential downside to offering this payment

is that workers may feel an obligation to work on our offered HITs and that could artificially

inflate the number of people who do at least one task and the number of HITs performed. This

should, however, not be a serious concern here since the new worker bonus does not vary across

the different conditions.

Figure 2: Letters to Prisoners Experiment—New Worker "Bonus"

Mechanical Turk allows requestors to require skills and "certifications" before workers can access their jobs. To achieve a sufficient sample size and because the jobs do not require any specific skills except for a decent command of English the only restriction we impose is that the computer accessing the HITs must be located in the US. This restriction also allows for an easier estimation of wage response of workers. It is in principle possible to circumvent this restriction through use of proxy servers, but if the computer appears to be located in the US, Amazon requires that workers provide a US tax ID number, which significantly limits the usefulness of using a proxy server to access Mechanical Turk.

5.1   Image Tagging Job

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, she is presented with five pictures per HIT and for each image is asked 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 **??** shows the page presented to a worker once she accepts the HIT, including one image.

The degree of agreeableness of the job is captured by the percentage of disagreeable images the worker is presented with. There are six levels in the experiment: 0, 20, 40, 60, 80, 100,

Figure 3: Image Tagging Experiment Page View

corresponding to 0, 1, 2, 3, 4, or 5 disagreeable pictures per HIT. The basic set of steps worker go through to complete the job is unchanged between the different levels of disagreeableness. The agreeable images cover a wide variety of topics such as garden pictures, nature, travel photo, food, and animals. We have a collection of 5921 of these pictures. The disagreeable images are identified using Google Image search terms and then false positives are deleted by us.[6] This process is, of course, open to cultural biases and influences in what is considered disagreeable but certain responses are more likely biological responses and those are the ones

---

[6] The Google Image search terms included topics such as amputations, autopsy, broken limbs, gangrene, gruesome injuries, larvas, vomit, and war atrocities to name a few.

we aim at.[7] The stock of disagreeable images consists of 1131 pictures. Once a worker accepts the image tagging job a randomized list of pictures is generated using the disagreeableness level assigned to the worker. The number of disagreeable pictures do not change between HITs, but the ordering is randomly allocated, so that a worker with, say, one disagreeable image per HIT (20 percent) may see that as the first image on one page and as the third on the next.

Figure 4: Image Tagging Experiment—Training and Test

The cost of learning is difficult to capture in a setting where the tasks themselves are relatively short and simple. The basic problem is designing the cost of learning condition such that we capture the difference in cost of learning without making the job itself easier or harder or otherwise fundamentally changing the job for workers exposed to different characteristics. We solved this by including a "training component" with or without a "test." Everybody was asked to read a description of different categories of tags and examples of each. Those selected for "training" condition got 15 questions to answer, where they were asked to categorize a set of

---

[7] Furthermore, the three authors are from Denmark, USA, and Czech Republic, respectively, and the images selected were images that all three of us agreed were disagreeable.

tags based on what they had just read. Workers could not go on until they had answered all correctly. Workers not selected for "training" were asked to click a button indicating that they had read and understood the content. Figure 4 shows the guidelines and the test questions.

As mentioned a potential issue with is that workers may see the task as changed because of the training and the associated test. Specifically, they may consider it more difficult once they have gone through the training because they think it reveals something about is required to get their tags approved. We can test whether the training changes the way the workers perform by examining how long they spend on the tasks under the different training conditions. If workers who have gone through the training spend more time coming up with tags, because they are more concerned about getting them right, we should observe longer time spent per HIT than those who have not gone through the training. We can also test for the occurrence of specific words or tags that we use in the training. We would expect those who have gone through the training to be more likely to use words that are used as examples in the training, which should make the job easier.

Figure 5: Image Tagging Experiment—Approval rate, pay, and availability
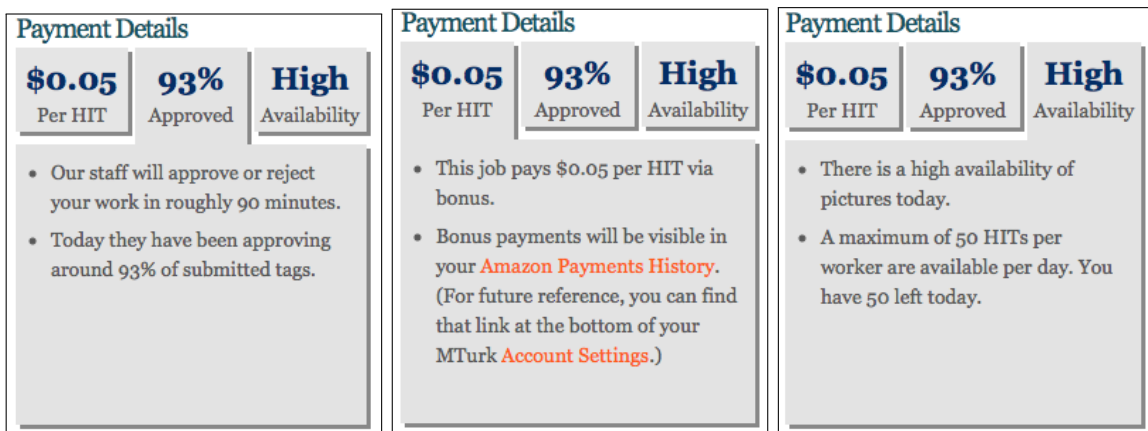
The probability of success is captured by our "approval" rate for tags. Either 93% or 56% were listed as approved. Figure 5 shows an example. We paid everybody for all work irrespectively of the assigned approval rate. Furthermore, we never rejected HITs. Rejecting HITs may have hurt workers since some requestors restrict access to job by requiring a certain acceptance rate. We did not explicitly say that we would not reject HITs, but some workers may have inferred

that we would given the vague phrasing for the text, which could mean either non-payment or outright rejection of HITs.[8] Given the potentially negative effect on future access to work it is likely that the effect of this condition is larger than simply the difference in expected pay.

For the image tagging experiment we did not directly test the availability, but simply indicated that there were high availability of images and that a worker could work on 50 HITs per day. The final part of the experiment is the pay offered. 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. Figure 5 shows an example of pay and availability.

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 would be assigned a new test of job condition. We choose 07.58 GMT because that was the time of the day where there were the fewest number of workers on Mechanical Turk.[9] 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.

5.2 Letter Writing Job

In the letter writing job the basic task is to write a positive and supportive letter to a prison inmate. We use different types of prisoner offenses to capture the agreeable/disagreeable aspect of the compensating wage differentials theory. One group writes letters to sex offenders and the other group writes letters to prisoners with a crime that could be perceived as less disagreeable. Figure 6 show the two conditions for agreeable/disagreeable.

Cost of learning is captured with a training component with or without a "test." Everybody was asked to read the guidelines. Those selected for "training" condition got two questions to answer. Workers could not go on until they had answered both correctly. Figure 7 shows the guidelines and one of the test questions. A potential issue with the training is that workers

---

[8] This is borne out by emails from workers addressing the approval rates.
[9] A worker working right around the change point could potentially see one set of condition initially and another set later on.

Figure 6: Letters to Prisoners Experiment—Agreeable vs Disagreeable

may see the task as changed because of the training and the associated test. Specifically, they may consider it more difficult once they have gone through the training because they think it reveals something about is required to get the letters approved. We can test whether the training changes the way the workers perform by examining how long they spend on the tasks under the different training conditions. If workers who have gone through the training spend more time writing the letters because they are more concerned about getting them right, we should observe longer time spent per HIT than those who have not gone through the training. We can also test for the occurrence of specific words that we use in the training. We would expect those who have gone through the training to be more likely to use words that are used as examples in the training.

The probability of success is shown by our "acceptance" rate for letters, although we pay everybody who submit acceptable letters. Either 94% or 51% were listed as accepted and Figure 8 shows an example of a high approval rate. Finally, constancy is modeled by varying the limit on the number of HITs available to the worker. Either 90 or 9 HITs were available. Figure 8 shows a low availability example.

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

15

Figure 7: Letters to Prisoners Experiment—Cost of Learning

Training Guidelines for Writing a Prisoner

You must learn these rules to write letters to prisoners. Your letter will be **rejected** if it:

1. Depicts or describes procedures for the brewing of alcoholic beverages, or the manufacture of drugs or other intoxicants;
2. Is written in code;
3. Depicts, describes or encourages activities which may lead to the use of physical violence or group disruption;
4. Encourages or instructs in the commission of criminal activity;
5. Is rude or disrespectful;
6. Contains racist, chauvinistic, sexist or obscene comments. Hopefully that needs no further explanation;
7. Discourages the prisoner from successfully continuing in his civil life upon release;
8. Lacks compassion and positive outlook;
9. Describes an illegal activity in a positive way leading to distortion in role model creating process on the side of the prisoner (e. g. describes a criminal as a hero so that the prisoner grows to admire him and hence starts to perceive criminal activity as attractive or laudable);
10. Undermines the correctionary process the prisoner is undergoing in the correction facility;
11. Contains copy/pasted text other otherwise is not earnestly written to the prisoner.

Testing Your Understanding of the Guidelines

Read each of these ficticious letters, and identify the rules they violate.

- "Dear Carlos,
  I cannot possibly comprehend how a human being might be capable of doing what you did. However, you were justly convicted and you are paying your debt to society. So I guess it would be wrongful of me to deny you my friendship. On the other hand, it is not quite clear whether you have been praying to God to ask you forgiveness and whether it will be granted to you. I guess it all depends on the entirety of your life, personality and actions. Anyway, I find difficult to relate to you as a person. Nevertheless, I would like to say that I hope that you won't deviate from the righteous path again and pray for God's forgiveness. Dabit deus his quoque finem.

  Yours Truly,
  Jacob"

List 4 of the above rules that this letter violates separated by commas (e. g. if it violates the first and second items, write "1,2" in the text field below).

Items Violated: [＿＿＿＿＿]

Figure 8: Letters to Prisoners Experiment—Probability of Success and Constancy

Write a Short Letter to an Inmate

Inmates need moral support from outside of the prison walls. Research shows that inmates with positive contacts outside of prison are less likely to return to prison, crime, and substance abuse, and more likely to find a job upon release.

Read the following prisoner's bio, and write a compassionate letter. Please do not include your email address, full name or address in the letter.

Marcus T.'s profile

Offense
CPC Section 288a(c): Oral Copulation, Victim Under 14/Defendant 10 Years Older or by Force, Violence, etc.

Bio
Hi, I'm Marcus J. T. and I'm from Portland, Maine. I'm 45 years old, 6'5 and weigh 220 pounds – big and muscular. I have dark eyes and hair. I love writing poems and listening to the classic jazz and soul greats. But there's something I'm still lacking even after all this time, and that's a genuine love with a woman where we can be forthcoming and respectful about our past mistakes or triumphs, our hopes and dreams, etc. I just need a chance to prove myself as somebody who's worth taking the time to trust with your heart. I have a bit of a formula for the sort of relationship I'm seeking. It's companionship-honesty-faithfulness-open to listening-and talking. If you'd like to write me, it would be great if you could send an up-to-date photo of yourself along with your response. I'm seeking women from their mid-20's to their mid-50's. Race doesn't matter to me.

Payment Details

$0.10	94%	9
Per HIT	Approved	Available

- Our staff will approve or reject your work in roughly 30 minutes.
- They are currently approving 94% of submissions.
- You must do the training before working.

Write a Short Letter to an Inmate

Inmates need moral support from outside of the prison walls. Research shows that inmates with positive contacts outside of prison are less likely to return to prison, crime, and substance abuse, and more likely to find a job upon release.

Read the following prisoner's bio, and write a compassionate letter. Please do not include your email address, full name or address in the letter.

Marcus T.'s profile

Offense
CPC Section 288a(c): Oral Copulation, Victim Under 14/Defendant 10 Years Older or by Force, Violence, etc.

Bio
Hi, I'm Marcus J. T. and I'm from Portland, Maine. I'm 45 years old, 6'5 and weigh 220 pounds – big and muscular. I have dark eyes and hair. I love writing poems and listening to the classic jazz and soul greats. But there's something I'm still lacking even after all this time, and that's a genuine love with a woman where we can be forthcoming and respectful about our past mistakes or triumphs, our hopes and dreams, etc. I just need a chance to prove myself as somebody who's worth taking the time to trust with your heart. I have a bit of a formula for the sort of relationship I'm seeking. It's companionship-honesty-faithfulness-open to listening-and talking. If you'd like to write me, it would be great if you could send an up-to-date photo of yourself along with your response. I'm seeking women from their mid-20's to their mid-50's. Race doesn't matter to me.

Payment Details

$0.10	94%	9
Per HIT	Approved	Available

- We have 9 HITs available for you today.
- You have completed 0/9 so far.
- The number of HITs available depends on the demands of the day.

experiment ran only through one 24 hour segment.

5.3 Estimation Strategy

The standard model for testing compensating wage differentials theory is regressing job and worker characteristics against the observed wage. We estimate a similar model here. The main difference is that we are not observing worker characteristics and that our observed "wage" is per job done rather than an hourly wage. For the sample of workers that perform at least one HIT we estimate

$$w_i = \alpha + \mathbf{c}_i \beta_1 + \epsilon_i,$$

16

where w_i is observed pay per HIT for worker i who have performed at least one HIT and $\mathbf{c}$ is a vector of variables capturing the different job characteristics. In addition, we estimate the determinants of total pay earned

$$totalpay_i = \alpha + \mathbf{c}_i\beta_1 + \epsilon_i,$$

for two samples: those who have performed at least one HIT and the complete sample.

If self-selection, broadly defined, is responsible for the lack of significant effects of job characteristics on wage, the sample of workers who perform one or more HITS should yield estimates for effects of job characteristics on wage and total pay that are small and statistically insignificant, while the full sample for total pay should yield statistically significant effects.[10]

One of the issues is that we do not observe a total pay different from zero for those people who never completed a HIT. We could either add a small number of the total pay or we can use the Heckman two-step method to control for selection. This is possible because we observe selection into work.

The experiments produce data on compensation per HIT performed, number of HITs completed, and dummies capturing which conditions a given worker was exposed to. The basic estimated equation is

$$Y_i = \alpha + \beta_1 w_i + \mathbf{c}_i\beta_2 + \epsilon_i, \tag{1}$$

where w is wage and $\mathbf{c}$ is a vector of variables capturing the four different job characteristics.

We estimate three models for each job. The first is OLS estimates of the effect of wage and job characteristics on the decision to work on the given job, which is the extensive margin. The second is OLS estimates of the effect of wage and job characteristics on the number of HITs performed given that the worker has already decided to do at least one HITs, which is the intensive margin. Finally, we estimate a censored regression model that takes into account the lower bound censoring that occur at zero HITs and the upper bound, which vary from job to

---

[10] Estimating this regression on the full sample would just yield a test of balance and no economic information.

job and depending on the job characteristics presented with.[11]

6 Results

6.1 Image Tagging

During the six 24 hour segments the image tagging experiment ran we has a total of 3,036 unique workers. Focusing here only on the first 24 hour segment a worker joined, 1,101 workers completed one or more HITs.

Table 1: Results for Image Tagging Experiment

	Workers with one or more HITs		Full sample
	Log wage	Log total Pay	Log total Pay
Disagreeableness	0.064	−0.220***	−0.147***
	(0.060)	(0.072)	(0.032)
Learning cost	−0.010	0.025	−0.111***
	(0.041)	(0.049)	(0.022)
Lower probability of success	−0.012	−0.082*	−0.104***
	(0.041)	(0.048)	(0.022)
Intercept	−1.434***	0.900***	0.460***
	(0.043)	(0.051)	(0.025)
Observations	1,101	1,101	3,036

Notes. * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers that have completed at least one HIT.
[b] Of the 837 observations, 546 were left-censored, 253 were uncensored, and 38 were right-censored.

Table 2 show the results for the three estimated models: extensive margin, intensive margin, and censored regression model. Not surprisingly there is a statistically significant and large effect of increasing wage on both the probability of a worker doing at least one HIT and the number of HITs performed if working.

What is especially of interest is that the three job characteristics, disagreeableness of the images, cost of learning, and the probability of success have substantial and statistically significant negative effects on the probability of working, but only the disagreeableness had a statistically

---

[11] In the case where there are only a fixed lower and a fixed upper bound censoring point the results from this estimation will be the same as that from a Tobit model.

Table 2: Results for Image Tagging Experiment

	Margin		Censored
	Extensive	Intensive	
	Probability of Accepting Job	Number of HITs Performed All who did one or more HITs[a]	
Log wage	0.059***	2.901***	0.187***
	(0.012)	(0.567)	(0.037)
Disagreeableness	−0.092***	−3.910***	−0.276***
	(0.025)	(1.136)	(0.074)
Learning cost	−0.155***	−0.073	−0.453***
	(0.017)	(0.770)	(0.051)
Lower probability of success	−0.095***	−0.993	−0.283***
	(0.017)	(0.761)	(0.050)
Intercept	0.622***	13.838***	0.473***
	(0.027)	(1.141)	(0.078)
Observations	3,036	1,101	3,036[b]

Notes. * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers that have completed at least one HIT.
[b] Of the 837 observations, 546 were left-censored, 253 were uncensored, and 38 were right-censored.

significant effect on the intensive margin.

6.2 Letter Writing

Table 3 show summary statistics for the 2111 workers observed. The workers were spread almost equally across the possible combinations. There are more workers who did no work when presented with the more offensive condition than when presented with the less offensive condition. This indicates that the compensating wage differentials theory is supported on the extensive margin.

Table 3: Summary Statistics for Letter Experiment

	Accept Job		Number of HITs	
	Better	Worse	Better	Worse
Agreeableness	0.311	0.238	7.059	8.185
Learning	0.333	0.220	7.582	7.514
Probability of success	0.300	0.250	6.694	8.507
Constancy	0.278	0.269	11.191	3.545

Table 5 shows the results for extensive and intensive margins for the letter writing task.

Agreeableness, learning, and probability of success all have strongly statistically significant effect on the probability of working on at least one HIT. Being presented with the more disagreeable prisoner type, higher cost of learning, and a lower probability of success all lead to fewer workers completing one or more HITs. There is, however, no effect of the job characteristics on the intensive margin and some of the effects even have the wrong sign. In addition, it appear that Mechanical Turk workers respond to economic incentives with both the likelihood of working and the average amount of work increasing with increasing pay.

Table 4: Results for Letter Experiment

	Workers with one or more HITs		Full sample
	Log wage	Log total Pay	Log total Pay
Disagreeableness	0.003	0.055	−0.068**
	(0.054)	(0.081)	(0.031)
Learning cost	−0.040	0.054	−0.111***
	(0.054)	(0.082)	(0.031)
Lower probability of success	−0.049	0.001	−0.056*
	(0.053)	(0.081)	(0.031)
Less constancy	0.051	−0.307***	−0.099***
	(0.053)	(0.081)	(0.031)
Intercept	−0.657***	1.218***	0.476***
	(0.057)	(0.086)	(0.036)
Observations	578	578	2,111

Notes. * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers that have completed at least one HIT.
[b] Sample consists of all workers that have completed at least one HIT and who were not presented with the low availability condition.
[c] Of the 2,111 observations, 1533 were left-censored, 510 were uncensored, and 68 were right-censored.

These results confirm that an important reason why prior research has generally not found support for the compensating wage differentials theory is because of self-selection. In our experimental set-up self-selection in off-line labor markets show up as the decision whether to accept a job or not.

6.3 Worker Fixed Effects Results

This section will cover the worker fixed effects results from the image tagging experiment.

Table 5: Results for Letter Experiment

	Margin			Censored
	Extensive	Intensive		
	Probability of Accepting Job	Number of HITs Performed		
		All who did one or more HITs[a]	High availability only[b]	
Log wage	0.073***	4.186***	6.650***	6.175***
	(0.014)	(0.981)	(1.790)	(0.934)
Disagreeableness	−0.074***	0.827	1.651	−3.717***
	(0.019)	(1.257)	(2.351)	(1.232)
Learning cost	−0.110***	0.116	−0.124	−5.789***
	(0.019)	(1.263)	(2.378)	(1.237)
Lower probability of success	−0.054***	1.440	2.327	−2.127*
	(0.019)	(1.253)	(2.344)	(1.228)
Less constancy	−0.014	−7.728***	.	−3.415***
	(0.019)	(1.253)	.	(1.228)
Intercept	0.462***	12.949***	13.938***	−3.012*
	(0.025)	(1.478)	(2.544)	(1.547)
Observations	2,111	578	303	2,111[c]

Notes. * sign. at 10%; ** sign. at 5%; *** sign. at 1%.
[a] Sample consists of all workers that have completed at least one HIT.
[b] Sample consists of all workers that have completed at least one HIT and who were not presented with the low availability condition.
[c] Of the 2,111 observations, 1533 were left-censored, 510 were uncensored, and 68 were right-censored.

7 Conclusion

[TO BE ADDED]

References

Brown, C. (1980): "Equalizing Differences in the Labor Market," *The Quarterly Journal of Economics*, 94(1), 113–134.

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.

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

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.*

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

Rosen, S. (1986): "The Theory of Equalizing Differences," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and R. Layard, vol. 1, chap. 12, pp. 641–692. Elsevier.

Shapiro, C., and J. E. Stiglitz (1985): "Equilibrium Unemployment as a Worker Discipline Device: Reply," *American Economic Review*, 75(4), 892–93.

Smith, A. (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations.* W. Strahan and T. Cadell.