Online Labor Markets I: Some Issues in Designing Online Labor Markets

March 1, 2017

Guest Speaker: John Horton, NYU Stern School of Business

1 John’s Background

- Completed a PhD in Public Policy from Harvard Kennedy School.
  - He did the doctoral program on Inequality and Social Policy.
  - John mostly identifies as an economist, but has published a bit in CS (EC, CSCW, HCOMP) – he took an early interest in crowdsourcing and human computation.

- He spent two years as a staff economist at oDesk (now Upwork) after doing observational work with their data – he joined to be able to do more interesting stuff.
  - Mostly as product manager, designing features to solve marketplace problems, figuring out what features works, were they making progress; he also designed lots of experiments.

- Started at NYU Stern in 2013 in the Information Systems group.
  - Spent summer 2015 and 2016 at Uber.
  - Spent summer 2016 at Airbnb (and half time with Uber).

2 What is Special About Computer-Mediated or Online Labor Markets

As a labor economist, we care most about how the initial match is formed. Is there statistical discrimination or taste-based discrimination? If workers are not finding jobs, it’s because of search frictions.

Historically, that part of the labor market is very poorly observed by the data that the bureau of labor statistics collects. They generally can see a person, their job, their wages, their hours, but not the matching process. Online markets’ real strength is that they can be used as a test bed for doing social science. You can see the entire process: the posted job opening, people applying with a wage bid, employers deciding who to interview, who they hire, at what terms, and how the contract progresses. In traditional markets, you get a tiny slice of that, and all after the match has been formed.
• **Question:** Is it assumed that the way that job matches form and any discrimination that you’d observe online is closely related to what you’d observe offline? Is there evidence that you can make comparisons like that?

**Answer:** As with any social science, you have to draw the distinction between internal validity and external validity, and external validity is always going to hinge on: do you think you’ve identified some mechanism that is likely to be very general? If I were to calculate the elasticity of labor demand on MTurk, who cares? It’s not the elasticity of labor demand in the market more generally. But if I can find in a large online labor market that employers tend to put a lot of weight on gender, except when they have some other information that indicates it was about statistical discrimination, that seems more likely to be general and not too context-dependent. One of the reasons for doing more work on something like Upwork instead of MTurk work is that Upwork looks more like a real labor market.

• **Question:** What is the difference between MTurk and a large online labor market?

**Answer:** MTurk’s tagline is “artificial artificial intelligence.” The tasks are the sort of thing you should be able to do with an API call: does this image contain a cat? Both sides are anonymous. There’s no selection; there’s no screening. The tasks are usually for pennies. You put your work out there, and someone works on it if they meet the qualifications. See [Are Online Labor Markets Spot Markets for Tasks?](#), a paper on how Turkers think about the work they’re doing, what expectations they form for employment. By comparison, as large online labor markets has people trying to hire for relatively longterm relationships: website design, data entry, bookkeeping – you’re going to know the workers personally, interview them, and go through a selection process. This is more like the traditional labor market.

• **Why computer-mediation in labor markets is a big deal:**

  – You can observe what you can’t typically observe in conventional markets.
    
    * Blog post on [the digitization of the supply side of the market](#)

  – You can also manipulate things you can’t typically manipulate.
    
    * [The Online Laboratory: Conducting Experiments in a Real Labor Market](#)

  – You can change how the market works in some fundamental way, which you cannot do in a conventional market. E.g. making it so the employer cannot access the wage history of the applicant. There’s been some attempt to regulate this in conventional markets, the concern being that employers either screen or bargain your wages down based on wage history. In John’s paper on [wage history](#), they ran an experiment where employers could not see the wage history, and they can also search the text of message interactions to see if it was asked for or exchanged. You could compare cities where this has been regulated to those who haven’t, but that might not be credible, and you can’t get into the micro-mechanisms. In their paper, you can actually see how it impacted screening. Employers who couldn’t access wage history expanded the number of applicants they interviewed, and ended up hiring people who had lower wage histories. The workers also seemed to strike a better bargain.

  – There is a lot of space for market design (more on this below).
Question: What can employers typically observe?
Answer: A profile picture, name, past jobs, feedback (bilateral reputation system like eBay/Airbnb). In the past, they could see: the applicant, the past job, that they worked for 20 hours and were paid $400. In the experiment, what they needed to calculate the average wage was stripped away. There’s a lot of policies intended to help people by removing information from the matching process, e.g., “ban the box” asking about criminal convictions, or taking out gender or race/ethnicity information from applications. The concern is that to the extent that the discrimination that occurs is statistical discrimination, e.g., using gender as a proxy for productivity, the worry is that removing it will lead to a cruder form of statistical discrimination. The concern with “ban the box” is that removing criminal history will only benefit white criminals, who actually would have been screened out, and then disadvantage other groups. This is an important inequality question, and in the online labor market setting, we can change the whole information set that the employer has access to. Then there’s a generalizability question: are they jobs like other jobs? Also: how would this work in equilibria? In the paper, we try to model how workers would adjust their bidding strategy if employers could not screen based on wage history.

• How online labor markets help overcome geography:

- The lack of labor mobility is probably the most consequential fact about the world from a global inequality standpoint. If you asked economists what the best thing that you could do for someone to raise their income, aside from transferring money to them directly, they would say letting them migrate. However, there’s not a lot of political enthusiasm in western countries for increased migration. Online markets are kind of a hack around this, letting people virtually migrate and work online. However, this hasn’t become as big of a deal as I’d originally hoped: work seems to be getting even more concentrated. (See debate on the geography of jobs.)

  * See “Economics and Emigration: Trillion Dollar Bills on the Sidewalk”

- Some work on how workers think about what they are doing:

  * “Are Online Labor Markets Spot Markets for Tasks?: A Field Experiment on the Behavioral Response to Wage Cuts” [4]
  * “The condition of the Turking class Are online employers fair and honest?”

Question: Are you suggesting that online labor markets are responsible for jobs getting even more concentrated, or that they haven’t had the counter-effect that you’d hoped?
Answer: They’re certainly not responsible, but they’re not important. It’s a nice-sized business, but it’s trivial compared to the size of the labor market. The hope that this would be a way for lots of people to access western labor markets does not seem to have played out. Information and communication technology is getting better all the time, but it still seems to be not as good as being physically colocated.

Question: Could that partially be a result of regulations in traditional labor markets in the US that don’t apply to, say, MTurk?
Answer: I’m not too familiar with what people get paid on MTurk. In a large online labor market, you can’t really find a place premium. People get paid about what their
skills will bear in the market. If there were a lot of onerous regulations in the US that prevented people from hiring as freely as they liked, it should make these markets more attractive. However perhaps they haven’t been as big of a deal because hiring someone through these platforms is still a bit of a pain: you have to manage people directly; there are time differences and cultural differences. Paying people based on where they’re from is the largest violation of the law of one price. In large online labor markets, people seem to get paid based on their observable skills, indicating that they are pretty competitive.

• Some more of John’s work on online labor markets as a phenomena:
  – WINE paper (from 2010) Online labor markets
  – NBER Chapter (from 2013) Digitization of the contract labor market
  – “Owning, Using and Renting: Some Simple Economics of Sharing Economy” (more about companies like Airbnb and Uber, but some discussion of what these computer-mediated markets need)

3 Design Challenges and Opportunities

• Recommendations:
  – Much of labor economics is about search frictions: the difficulties of knowing what jobs have opening and how much they pay. On the demand side, this includes knowing who’s on the supply side and what skills they have. This lack of information is almost the designing characteristic of labor economics. Online, we can do all of the recommender system tricks: people who like A bought B; people who viewed A viewed B. This can help to overcome the search frictions. But it’s more strategic: perhaps people care more about you recommending them as a person than products do.
  – John worked on using the application graph to suggest workers who apply for similar jobs, exposing similar workers on their own profiles. Unsurprisingly, the workers hated this, as it was advertising their competition.
  – Question: Products (or the people who work for them) do care if you advertise their competition.
    Answer: Consider paid search vs. organic search. Google leaves organic search alone. On Amazon, this is like leaving the rankings of which toasters did better alone, and brands accept this. However, in the context of a labor market, a ranking creates more anger because it’s so personal and consequential to someone’s livelihood. This shows up a lot in reputations. In Yelp data, reputations are centered at 3.5 stars. In large online labor markets, reputations are incredibly right-skewed – almost everyone gets five stars, and if you give someone a bad rating, they get really upset and they’ll complain or try to offer refunds. Humans are fine saying something bad about a product, but reluctant to do so about another person.
  – One concern with using recommendations is crowding out workers. Certainly they can help A do better than B, but that can’t be the standard from the platform’s
perspective. To what extent does B get crowded out or harmed? See John’s paper about an experiment where they made recommendations (forthcoming at JOLE): [The Effects of Algorithmic Labor Market Recommendations](#). Crowd-out effects were not really detectable. In this case, likely because a lot of employers don’t fill their jobs anyway – the fill rate is only 50%. As the fill rate gets close to 100%, recommendations just move around who gets hired. However, at 50%, it just expands who gets hired without crowding out.

- **Question:** Is the effect different than in marketplaces like Amazon?
  
  **Answer:** There’s a large literature on recommender systems: how to design and build, and also some empirical work. [Kartik Hosanagar](#) at Wharton has a lot of work on introducing recommendation systems with experiments. I’m sure that in those cases, there’s a fair amount of crowd-out. However, in equilibrium, does it raise purchases? My understanding is yes, but it’s probably not pure-substitution, but rather somewhat market-expanding.

  **Comment:** In the 90s literature on websearch, there was the heavy-tailed distribution of popularity, and they would ask the question: does algorithmic search make the heavy tail even heavier by ranking the popular pages more highly and driving traffic to them, so the rich get richer?

  **Answer:** This sounds similar.

- **Question:** If I buy something on Amazon, I just want it to work, so if Amazon is willing to recommend it, it’s probably not going to break on arrival. I’m not looking to buy the *best* toaster. If I’m hiring someone to design a webpage, I want the best webpage, so the platform’s recommendation might carry less weight to me.

  **Answer:** The recommendation was: we see you’ve posted this job, you should consider recruiting these people. Not: from an applicant pool, we recommend this one to the other. There are employers who just want something done well enough and are very price-sensitive, and there are people who want something exceptional.

- There is a persistent and justified worry that algorithms will just learn about statistical discrimination that is already in the data. (John’s paper doesn’t deal with this.)

* **Employer and Worker Reputations:**

  - Reputation systems have serious flaws, but we shouldn’t throw out the baby with the bath water.

    * [Blog post on this](#)

  - There are issues, however:

    * Bilateral reputation systems in online labor markets seem to quite prone to “reputation inflation.” The distribution is highly right-skewed.
    * Reputations which are private, then aggregated, then made public seems to “work,” but there are fairness issues.
    * John has a paper on this with Joe Golden and Apostolos Filippas that is currently being revised – here’s the “old” version: [http://john-joseph-horton.com/papers/private_feedback.pdf](http://john-joseph-horton.com/papers/private_feedback.pdf)
* John would love to see more work on eliciting feedback and using it more constructively (particularly feedback that was focused on helping sellers improve rather than just help buyers select).

- **Incentivizing sides of the market to truthfully conveying preferences or attributes**
  - price-sensitivity [13] and workload capacity [8]:

  - **Revealing preference in the price-sensitivity vs. quality tradeoff**: Sometimes it’s not strategic to be truthful about what the agent cares about, i.e. to say a web programming job is really an accounting job. For an employer to say, “Price is no object. I’m not price-sensitive and I have a virtually unlimited budget. I just want the best product.” Workers might bid up on you, knowing that you’re not price-sensitive. If you say you’re very price-sensitive, you might save money and people might shade down their bids, but you might get really bad applicants. John has a *paper* with Ramesh Johari [13] on “relative preferences over pricing quality” – an experiment where employers were asked if they were price-sensitive and were told it might be revealed to the market. The platform randomized whether this was revealed to workers or not. From this, we can observe to what extent the workers sort, and how much the workers bid up on the employers. We looked over time to see if employers liked having the signal; the conclusion is that it seems to give people the information they want to find the right trading partner. Also, in the bigger picture, this information feeds back into the algorithmic recommendations. In one platform, they try training predictive models on the text of posted jobs, which is unstructured data, but they seem to get a lot more mileage out of identifying the dimensions that are important to either side of the market and training on more structured data, if you can get people to tell the truth approximately. In conclusion, the price-quality tradeoff information was useful itself and also in recommendations.

  - **Revealing capacity to take on more work:**
From “Buyer Uncertainty About Seller Capacity” [K]:

Figure 1: Distribution and acceptance of recruiting invitations by worker weekly hours-worked bins

The top panel is how many invitations workers get from employers. The x-axis is hours worked per week percentile. The extreme left is people not working much; the extreme right is people working a ton of hours. The workers working a ton of hours get a ton of invitations, and if you were to break this down further, you would see that it’s even more extreme. The 95th percentile to the top are getting over ten invitations per week, whereas almost everyone else in the market is not getting very many at all. You can see that the invitation acceptance rate plummets as you go higher. There’s a whole bunch of reasons that this could just be selection and employers could be rational about it, but this paper is about how it’s probably not. If you gave employers more information about worker capacity, they would adjust their invitations and invite people who actually have the capacity to take on more.

The mechanism design aspect of it (not pursued in this paper) is that workers don’t have a strong incentive to tell the truth here. If offers are valuable even if not pursued (e.g. academic job offers, which can be used in negotiation), employer-recruiting-attention is a big source of inequality in the market: the rich get richer. If there were ways to get workers to be more truthful about their capacity to take on more projects, it would be very useful in the market. In conventional markets, one of the reasons we don’t see employers pursuing workers as much as the reverse is that employers have a hard time conditioning on who is willing to leave a job, so they have to wait for applicants to come to them. LinkedIn is trying to take advantage this by adding a new feature where you can privately signal that you’re actually looking for work, but people are reluctant to let their current employer know that they’ve got one foot out the door.

- This seems like a great market design area – getting people to tell the truth about things they have at least some incentive to lie about is arguably the beating heart of
mechanism design.

- **Shortcutting Statistical Discrimination** – eliminating employer access to wage history:
  - Huge problem in all labor markets:
    * Taste-based discrimination
    * Statistical discrimination
  - For taste-based, we can just eliminate that “signal” from the process; statistical discrimination is a bit more fraught.
    * Will they stop hiring altogether?
    * Will they substitute towards “cruder,” less fair forms of discrimination?
  - John’s paper on wage history with Moshe Barach.
  - Statistical discrimination is a huge problem in all labor markets. It’s a huge contributor to inequality. Not a lot of hiring is necessarily happening in online labor markets. If you think of hiring becoming more computer-mediated, where a firm uses some applicant-tracking system to collect resumes and present them to the employer, there’s going to be a lot more opportunities to design what kinds of information are allowed or not allowed in the process. Either voluntarily, if the firm cares about a just hiring process, or by statute, making it illegal to ask or condition on certain things. This paper described earlier was about using that design power to not allow employers to condition on wage history. The platform was concerned that without access to wage history, the firm might not hire anyone at all. That didn’t appear to be the case. Hiring actually went up a bit, although that might have more been a result of the fact that you can end contracts quickly here, so you can think of hiring as a continuation of the screening process. More information might just talk you out of hiring someone.

- **Question:** Is it common in various online labor markets that contracts aren’t a very large commitment; you can end contracts easily and use it as part of the screening process?
  - **Answer:** Like employment more generally, it’s completely at-will. There are no real protections here. Hiring is fairly low consequence. If it was guaranteed that once you hire a person, they have to work on your project until complete or for six months, I could imagine that if you took away information, you might get less information. At least in this experiment, that’s not what happened. The so-what of the paper was that we were able to induce the employer to do more evaluation. They looked at more candidates, interviewed them more intensively, and the text exchanges suggest that they were asking candidates more questions about their productivity. If we could get employers to substitute towards hiring less experienced workers, not relying on cheap signals that perpetuate statistical discrimination, but rather dig in a bit more and find out if the person is productive, that’s quite positive. There’s quite a bit of thinking in labor markets that there are suboptimal levels of investigation of people’s productivity, that there are spillover effects that don’t get captured from that kind of screening and evaluation. Pallais has a paper where she hired a bunch of people at random on Upwork. The treated workers went on to earn a lot more money in the future. They
got a benefit from this reputational signal of getting hired by someone else. Essentially, her results speak to the idea that there's too little of this kind of information that's getting out there into the market about who is good.

- **Setting Pricing Policies** – an experiment in imposing a minimum wage [10]:

  - Probably the most direct inequality-reducing or -enhancing policy is directly changing prices, but there are concerns about unintended consequences.
  - John has a [paper](http://example.com) and [presentation](http://example.com) on experimentally imposing a minimum wage in a large online labor market [10].
  - The price structure is to take a percentage of the wage bill, so there historically hasn’t been membership fees or a two-part tariff. The one exception to that was imposing a minimum wage. Before they did it, the big questions were:
    * If we impose a minimum wage on employers, are we going to just screen a lot of projects out of the market?
    * To what extent are employers just going to substitute toward other workers when they face the minimum wage? That paper is about that experiment.
  - The experimental portion: they took some employers, randomized a minimum wage for them, and did not allow the applicants to that job to bid below the assigned minimum. The status quo in the market was always no minimum wage, and maintained that for a control group.
  - What happened:
    * Almost no reduction in hiring, even under quite high minimum wages.
    * There was a large amount of labor substitution. When facing a minimum wage, employers hired more productive workers (as opposed to lower quality for cheaper). The number of hours worked per filled projects went down substantially. This could be due to workers scaling back their projects endogenously facing a hiring labor cost, but about half of the effects seem to be from hiring a more productive person.
  - How would this hold up in equilibrium? Given that if most employers try to adjust by hiring the most productive workers, you might be afraid that this adjustment wouldn’t work; it would just bid up the slightly more productive workers.
  - **Question:** More productive workers were also working more, is that accurate?
    **Answer:** Yes. The concern is that to the extent that they’re not fully extent, if everyone is trying to adjust this way, in equilibrium, you’re going to bid up their wages, and maybe this adjustment strategy gets foreclosed.
  - A minimum wage on everyone was one of the treatment cells. In the paper, they do differences around that imposition to see how much does this seems to happen in equilibrium. The experimental results are quite similar to the equilibrium to the results, but one caveat is that there’s some evidence that employers stopped posting the kinds of jobs that historically paid below the minimum wage, so it does look like there’s a shift in the composition of what jobs are getting posted. The other thing: we can look and see what happens to individual workers. The results are troubling: the workers at the lowest end of the market have to bid up a lot to comply with the minimum wage, and they get squeezed – their win probability per application plummets, and they either leave the market or are a lot less successful than they had been in the past.
Question: The minimum wage was enforced as a per-hour minimum wage? Is it the case that the less productive workers had lower quality output, or they just take longer?
Answer: It’s a mixture of both. In some categories of work, if you regress someone’s log keystrokes per minute on their wage, you can explain a lot of the variance; it’s literally their productivity. That’s not true in other categories (like programming). The kind of work we’re talking about which pays $2-4 per hour tends to not be highly conceptual stuff, is typically doing data entry where physical productivity and monetary productivity are tightly linked.

Question: Is it feasible for workers, if they know they’re only working at a $4/hour rate, to misreport working one hour when they worked two?
Answer: In this market, workers have to install time-tracking software on their computers that serves as a digital punch-clock, so it’s not really possible with the digital hours to shade your hours. That being said, you could potentially offer rebates, but we didn’t see any evidence that workers were doing that. Another adjustment strategy would be to shift to fixed-price work instead of hourly. We didn’t see much evidence of employers adjusting in this way — either employers haven’t thought this through, or there’s a pretty big cost to changing the contractual form, but we’re not sure exactly why we’re not seeing it. There’s also competitor platforms, so the minimum wage could shift work to other platforms. In the experiment, they were actually able to see some data about platforms, and there wasn’t any evidence of this. That is at least in part because employers still hired regardless, so there were no missing job openings that were posted on the other platform. Longer term, perhaps employers would react more strongly. The point of the paper is to measure the substitution effect when you can observe these past measures of worker productivity.

Question: How does training come in? In the traditional labor market, an employer can train you to be more productive. Are there avenues for training here if you’re working at a low rate and want to be more productive?
Answer: It’s a bit of a puzzle why firms train workers. It basically implies some kind of market power, because if the firm can make you more productive on their own dime, you’ll just go work for some other firm that didn’t have to pay you. The fact that some firms do offer general skills training is a positive thing, but it’s indicative of some friction. In this case, I would love to be empirical about this. I have a hypothesis that workers offer discounts to employers to work on technologies that are up-and-coming that they would like to get more skills on, so I would love to see a swift discount that they offer until they get trained, or conversely, like a Flash penalty — if you make me use Flash, I’m going to charge you more because it’s a dying technology. To the extent that this stuff happens, it’s not too formal (not sending you through Engineering bootcamp), but more in the form of offering discounts to train. This goes back to the question on minimum wage: if you pull up the ladder by foreclosing job opportunities for new workers, it gets harder for them. People have had this concern about minimum wages in conventional labor markets, and there are sometimes laws that will allow firms to pay below minimum wages if they offer training, or if the person is a teenager or an intern or something. Employers don’t typically make use of these laws, I’m not sure why — possibly due to inequality concerns within the firm. But people have thought about and tried to address limitations of minimum wage policies.
Question: When you talked about trying to post a per-contract minimum wage, what seems challenging about that is: as a worker, I know how much I value my time per job, but if I look at a job and it could take anywhere from 30-60 hours, it’s hard for me to gauge how much I’m willing to accept for it. If you have the option of posting a fixed per-unit price, the downside is that it’ll rule out certain workers. Or, you can post a vague per-unit price, which is good because it makes more workers eligible, but it is bad because no one knows what the per unit price is.

Answer: I think that there’s a reason why time-based compensation is so commonplace. Partly, because firms don’t like giving workers high-power incentives when they’re supposed to be collaborating/cooperating with each other, but also there’s this paper by Pat Bajari and Steve Tadellis [2] that talks about the cost-plus (hourly, in this context) vs. fixed-price contracts. In fixed-price, you get high-powered incentives for cost reduction, but the downside is that you have to spend a lot more time specifying the contract, writing out explicitly what you want done. To the extent that you don’t do that, you create a winner’s curse dynamic – the lowest bid is going to be the person who misinterpreted the requirements of the project the most. The preference in this market for hourly projects is happening because fixed-price contracts are really fraught in a lot of ways. If you know exactly what you want done, the incentivizing power of it is really awesome and you don’t have to monitor work, but hourly contracts are commonplace because most projects don’t fit that pattern.

Question: Do you know of theoretical work looking at this?

Answer: Other than that paper? It’s well-cited, so there’s probably a big literature that followed it. It’s pitched as an Industrial Organization paper, not mechanism design.

Question: If any of us want to do experiments, are there any interesting datasets that you could point us to, or do we basically need to go work for a platform?

Answer: It’s a good project for a grad student or first/second year professor to pitch to a company to come visit for a summer. What’s worked for me is being reluctant to talk about what research project I want to work on on day one. If there’s a company that’s interesting, collecting a lot of data, but isn’t so big that it’s gotten bureaucratic, if you can get a grad student in there who ideally would feel comfortable taking on a product manager type of role, that would be ideal. In the beginning in my experience, I started sending around memos about smart features, but I got no response. What I had to do was start designing the features myself, getting designers and engineers interested, pushing for things to get built, and getting the features in the product queue. This is time-consuming, but if the question is big enough and the payoff is high enough, it can be worth it. There are a lot of economists who are interested in the question of minimum wage, and to be able to do this experiment with high resolution data is probably worth a year of research time to be able to do.

References


