Labor Economists Get Their Microscope:
Big Data and Labor Market Analysis

John J. Horton, Ph.D. and Prasanna Tambe, Ph.D.[1]
NYU Stern School of Business

[1] Corresponding author: Email ptambe@stern.nyu.edu. Prasanna Tambe is very grateful to the Alfred P. Sloan Foundation for financial assistance. Portions of this article were modified from a presentation given to attendees of the 2015 NBER Winter Digitization Conference.

Abstract

This article describes how the fine-grained data being collected by Internet labor market intermediaries, such as employment websites, online labor markets, and knowledge discussion boards, are providing new research opportunities and directions for the empirical analysis of labor market activity. After discussing these data sources, we examine some of the research opportunities they have created, highlight some examples of existing work that already use these new data sources, and enumerate the challenges associated with the use of these corporate data sources.

Big Data and Labor Economics

Economics, by many accounts, is experiencing a data revolution. The emergence of new Internet data sources is transforming both the scale and granularity at which researchers can examine economic phenomena, ranging from e-commerce transactions to online search behavior to consumer decision-making.[1] Recently, some of these new data sources have enabled social scientists to study new aspects of labor market activity that have historically been difficult to analyze. Although some aspects of the "wiring" of the labor market have been considered in detail in the context of falling costs of communication,[2] the capture of massive volumes of fine-grained data on labor activity and its analysis have many new implications, especially for research into labor-based phenomena.

The digitization of these processes presents a significant opportunity because labor economics, perhaps especially so among the various branches of economics, relies heavily on administratively collected data sources. These data sources, such as the National Longitudinal Study of Youth (NLSY), the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), or the Job Openings and Labor Turnover Survey (JOLTS) are notable for their breadth and their quality. They are however, expensive to generate and therefore only infrequently collected and limited in their sampling and in the scope of the questions that they can be used to answer. Along some dimensions, including both the granularity of observation as well as sample sizes, the gaps in these administrative data sources can be filled by new data sources that are being generated daily by the activity conducted through Internet labor market intermediaries.

This article outlines how these Internet data sources are opening new avenues for research in the analysis of labor market activity. Specifically, it 1) describes some of the data sources that are emerging, 2) examines the opportunities that these data sources present to researchers in the context of early work that is already leveraging these data sources, and 3) discusses some of the potentially significant pitfalls that can confront researchers who would like to use these data sources for research, including sampling issues and operational challenges relating to obtaining access to corporate databases.

Overview of Data Sources

Labor markets have long been characterized by information asymmetries on both sides of the market. Would-be employers are uncertain about which workers are available and the attributes of these workers; Would-be workers are uncertain about whether potential firms would make good employers (or even which firms are hiring). These asymmetries have created entrepreneurial opportunities with the rising penetration of the Internet, such as online job boards, employer review sites, fully online labor markets and websites that, while created for other purposes, collect labor-market relevant information as a by-product.

Employment Websites

Professional job and employment sites, such as LinkedIn, CareerBuilder, and Monster, are among the oldest of these labor market intermediaries, having become popular during the late 1990's. These also include websites that serve niche markets, such as Dice.com, which is a job board oriented specifically towards IT workers, as well as a growing number of international job boards, such as Naukri or Zhaopin. Job boards are arguably the largest websites in terms of the populations they cover. LinkedIn, which is both a job board and a largely successful attempt to digitize the supply side of the labor market, reaches almost 300 million users globally. These sites contain information about a substantial fraction of the US workforce in its databases.

Job sites capture a number of different types of valuable information that are useful for labor market research, including a) workers' prior employment histories, b) employers' job openings, and c) about the matching activity that takes place to connect these two sides of the market.

Much of the data available through the employment histories that workers post on job boards, or resumes, would not otherwise be available through government surveys. Matched employer-employee data sets, such as the LEHD (Longitudinal Employer-Household Dynamics) data, contain a great deal of useful information on firms and workers but they are expensive to manage and access to important fields can be restricted. Employment histories generated through job boards contain rich information about prior employers, job titles, skills, and dates of exit and entry, as well as details about education. They also occasionally contain information about geographic mobility (i.e. city to city moves).

Job listings contain information about offered wages, the skills that are required to fill a particular job opening, and other details valuable for understanding labor demand as well as perhaps related to broader questions around unemployment trends. In addition to the job sites mentioned above, aggregate data on job openings across employment sites and corporate web sites is collected by companies like Burning Glass Technologies and Indeed.com. This aggregation is useful for understanding broader job opening patterns that are not well represented by looking at data from a single jobs site.

Finally, because employees connect with employers through these websites and vice versa, a wealth of information is generated about job search activity. For example, job boards collect information about which applications are viewed and by whom, how these views are converted into applications, and how application level variables affect the likelihood of receiving an interview request. Data are also often collected on how various aspects of job search affect outcomes. For instance, platforms can alter the way in which workers and firms are directed towards one another during the search process, such as by implementing algorithms to direct workers with particular skills and backgrounds to particular employers.

Online labor markets

There is growing interest in how platforms over which work can be transacted online, such as Uber, Lyft, TaskRabbit, UpWork, and Amazon Mechanical Turk, impact workers and career paths. The data generated by these platforms, therefore, is becoming important for understanding how work is transacted in the new economy, how workers build reputation, and how reputation matters for work outcomes.

Because these platforms are self-contained markets, they offer several unique advantages for research, including the ability to conduct experiments. One advantage of working with these data sources is the granularity with which they capture transactional data. Work that is transacted through these sources is recorded at incredibly fine levels of detail, including the sequence of jobs that workers accept, measures of performance for each job, bid offers, and even more fine-grained measures of worker productivity such as keystrokes, in the case of UpWork or similar intermediaries, or driver routes and speeds, in the case of Uber and Lyft.

These platforms are also promising because they often offer opportunities to take advantage of experimental designs, either "natural" experiments than can affect behavior in these markets (such as changes in oil prices or currency fluctuations) or field experiments, in which the researcher manipulates aspects of the platform to test how the market responds.

Work-Related Collaboration Technologies

Platforms for collaboration, such as GitHub, Slack, and Sourceforge, provide perspectives on the activities required to produce new software.

These sites provide detailed information about individual, time-stamped contributions that users make to different technologies, how users identify software bugs, how they assign software bugs to others, how they form networks of contributors, and how they choose which technologies to work on to maximize either signaling or the acquisition of human capital. In other words, these collaboration platforms collect detailed data on how labor is organized to produce software, at an extremely detailed level that has previously only been within the purview of the internal records of software firms. The fact that this collaboration occurs across firm boundaries has additional implications that are of interest to social scientists.

These websites also often contain human capital or career information about individual contributors, which can enable analysis of how contribution to these sites can impact short-run labor outcomes or career trajectories.

Knowledge Sharing Platforms

Knowledge sharing platforms, such as StackExchange and Quora, are examples of forums where workers ask and answer questions that are (potentially) related to work objectives. For example, StackOverflow is an active community dedicated to questions and answers related to technical questions, such as those dealing with software bugs or database design.

From a research standpoint, these platforms provide detailed information about how workers exchange the knowledge they require for new technical activities. The data are often tagged according to their relevance to a wide range of technological activities and topics, including different programming languages and technologies (e.g. C++ or Python).

Given the visibility of these forums, there are numerous questions that can be studied using the data from these forums, including those relating to how workers acquire new knowledge. Moreover, like the collaboration technologies discussed in the prior section, contributors generally receive point-based rewards for participating on these forums rather than monetary compensation which raises interesting questions about why workers choose to contribute to these discussion boards.

To answer such questions, these data sources can be combined with supplementary data on the contributors, such as their employers or prior career information which enables examination of questions that relate to how the worker's contribution patterns are influenced by their employers or how their contribution to these discussion boards impacts their career trajectories, either directly through the signaling that occurs through such platforms or by the human capital development that occurs by being active on these forums.

Social Media and Search Platforms

Social media platforms, such as Facebook and Twitter, as well as search platforms such as Google often generate data that can be used to inform questions related to job activity.

Users of these platforms often post information or launch queries related to job changes or job search activities. For example, users may search for information on job openings in particular cities, or they may tweet when they are looking for new jobs or when they have just found a new job.

These data sources are not as singularly focused on career objectives as the other data sources described here, but they have the advantage that they are extremely widely used, and that they can often allow what is essentially the development of real-time indices of job activity. They also often capture early signals of labor market activity, such as rising interest in skills or leaving one's job, that can provide useful indicators well before they can be captured in official statistics.

Education Platforms

Online instruction platforms such as Udemy, Coursera, Udacity, Khan Academy, or Smarterer collect very detailed information on how workers acquire new skills. The granularity of the learning interactions that take place on the website are potentially useful for learning about who selects into training, what they choose to learn, how they perform,and what we can do to improve the process by which workers learn new skills, which is likely to be increasingly important given the growing importance of continuous skill acquisition in a technology based economy.

Career Intelligence Websites

A final class of data platforms that are useful for labor market analysis are career intelligence websites such as Glassdoor, that collect information on how employees view their current and former employers. These sites collect information on what employees think of management, company culture, values, advancement opportunities, and other employer attributes. Moreover, users enter text reviews that contain information that can be used to derive specific measures of various aspects of company culture, benefits, and attributes that matter for understanding the employer-employee relationship.

Opportunities and Examples

Collectively, these intermediaries provide a wealth of data with which to analyze questions about labor market activity that have never before been possible. These data sets offer specific unique advantages that we discuss in detail below. They allow researchers to measure the previously unmeasurable---often bringing this new data to bear on old questions. They also allow the most credible empirical method to be brought to research questions, namely the randomized controlled trial. These experiments---as well as the samples used in observational studies---can be enormous, given the low cost of collecting and storing data in these markets. Online environments also introduce new kinds of information into labor markets---prominent examples being algorithmic recommendations or worker reputation.

Measuring the previously unmeasurable

These data sources offer a number of advantages in terms of granularity, often enabling measurement of phenomena that have been difficult or impossible to measure using prior data sets. Examples of this are the detailed skills data collected by LinkedIn, or the reviews of employer culture collected by Glassdoor. At a more “nano-scale” level, online labor markets such as oDesk (now UpWork) enable detailed tracking of how intensively hired contractors are working, down to the level of individual keystrokes for some projects.

Administrative datasets are often constrained by the cost of data collection: when each subject must be interviewed in person, there are few economies of scale in data collection. In contrast, the cost of data collection on platforms is almost entirely a fixed cost. If the platform is large, it can easily and nearly costlessly collect enormous amounts of data. This has a number of advantages. First, precision is improved, making parameter estimates more precise. Second, it becomes possible to select constructed samples that meet some special requirements for doing causal research---for example, observations around some discontinuity. Third, the robustness of results can be examined by looking at different sub-samples based on demographics, geography and so on without the penalties in precision that are normally associated when doing this with smaller samples.

In a program of research using CareerBuilder data, Marinescu and co-authors make extensive use of the fact they can observe the "application graph" of workers applying to job openings. Marinescu and Rathelot (2015) quantifies the effects of geography on worker application direction, concluding that while geography matters (workers dislike applying for jobs far from where they live), the effects are too small to explain much of the frictional unemployment of the labor market.[3] Measuring the geography of all job applications would have previously been impossible. Marinescu and Wolthoff (2015) use the CareerBuilder data to explore how textual characteristics (such as the language used in the job title) affect the matching process.[4] They find that the language of the job title can explain more than 80% of the variation in the education and experience level of applicants.

On the other side of the market, Tambe and Hitt (2013) use the information on workers' employment histories that is captured on CareerBuilder resumes to measure the flow of workers between organizations, in essence building a network graph of labor flows among organizations.[5] They use these data for IT workers to quantify the economic impact of spillovers from IT investment that are generated through the labor market. Ge et al (2014) use LinkedIn data on the employment histories of scientists to examine the robustness of patent information for tracking the labor mobility of scientists.[6]

These data sources can also be used to develop measures of human capital that can be useful for measuring firm-level differences in production activities, such as in levels of computerization. For example, Tambe and Hitt (2012) use employment history data to generate measures of firms' investments in IT labor over two decades,[7] and using data from the LinkedIn skills database, Tambe (2014) measures employers' investments in the human capital associated specifically with big data technologies.[8]

A non job board example is Fradkin and Baker (2015), who use Google Search data to construct a "Google Job Search Index" and then show that expansions in unemployment insurance decreased job search activity---a key policy question in the design of unemployment insurance policy.[9]

Several papers explore otherwise-unobservable phenomena in the context of oDesk. Ghani et al. (2014) using oDesk data, trace out the continued importance of ethnic similarity in outsourcing relationships---connections that would go unmeasured outside the context of oDesk.[10] Horton (2015) measures employer recruiting attempts on oDesk to measure the effects of spurned invitations on subsequent match-formation.[11]

Experiments

One exciting aspect of digital markets is that experiments are often simple and low cost. By their nature, all interactions between users, each other, and the platform are computer-mediated and these interfaces are relatively easy to modify. Further, on most platforms, the infrastructure needed for experimentation already exists: many companies "roll out" features experimentally as a precaution against bugs affecting too many users are once. They also collect copious amounts of data about what is happening on the site, both as a by-product of the functioning of the site as well as for analytic purposes. As the infrastructure for controlled experimentation and the instrumentation for collecting data are already built, experiments can be easy to conduct.

Some experiments can be run by the platform itself. Horton (2015) describes an experiment in which oDesk introduced algorithmic recommendations to employers about which workers to hire.[12] This intervention substantially increased hiring for technical categories of work. Horton and Johari (2015) present results from another oDesk experiment in which employers were asked for their price/quality preferences before posting their job openings---these preferences were then exposed to would-be workers, inducing substantial sorting by workers and potentially better matches.[13]

A non-oDesk example comes from Gee (2015), who conducted an experiment on the career-focused social networking site LinkedIn, where the experimental manipulation changed whether or not other applicants could see the count of other workers who applied to the job of potential interest.[14] In addition to highlighting the power to reveal new sources of information in online settings, the paper is remarkable for the sample size: it was a 2.3 million person field experiment.

Some experiments are in the "Experimenter as Employer" framework, where the researcher poses as the employer. This is true of Pallais and Sands (2015)[15] and Pallais (2014)[16] where the researchers hired workers to conduct data entry tasks. In Pallais (2014), the main finding was that workers experimentally given a first job where far more likely to be hired by subsequent employers---highlighting the importance of on-platform experience.

New signals

In addition to measuring the previously unmeasurable, online platforms often use their unique position to collect and make available data that no single market participant could previously access. For example, a common corporate use case for LinkedIn is using it to explore the graph of current employees for recruiting purposes. This elucidation of previously quasi-invisibile cross-company connections is something only platform such as LinkedIn can accomplish, given its ubiquity in certain industries. Li (2015) uses the network of employer viewing behavior on LinkedIn to examine how the peer grouping indicated by LinkedIn search activity affects can be used to explain financial performance.[17]

Using data from oDesk, Kokkodis and Ipeirotis (2015) show that reputation scores (i.e., the familiar five star feedback) can be substantially improved in their informativeness when a model is used in which the nature of the work for which the feedback was earned is considered.[18] This paper is interesting in that it offers an improvement over something that is commonplace in online markets---the reputation system---which is not even present in conventional markets. Horton (2015), also using the oDesk context, finds that exposing aggregated and anonymized "private" feedback about worker performance to future employers can substantially affect employers' decisions about whom to hire.[19]

New types of work/crowdsourcing

The rise of online markets is creating new kinds of economic interactions. For example, MTurk is arguably the largest spot market for labor in the world. It is also alone in that the labor relationships being intermediated often last only minutes and pay pennies. However, in the same way that frictionless planes and vacuums are ideal for studying questions in physics, "stripped down" markets that are simpler than conventional markets have research advantages.

The most comprehensive exploration of new ways of organizing production comes from the "human computation" community within computer science. Researchers largely focus on what kinds of new systems can be built by combining human intelligence and algorithms/machine learning. An example in Bernstein et al.'s (2012) "Soylent" which uses workers from MTurk to create a "human powered" editor and writing assistant "inside" MS Word.[20]

In the social science realm, there have been some explorations of how new kinds of labor market institutions and organizations affect outcomes. For example, Stanton and Thomas (2014) show that "agencies"---a kind of quasi-firm that vouches for worker quality at the start of their careers in online labor markets---provide a useful function.[21]

Challenges in Using Corporate Data Sources

Although there are, as described above, numerous opportunities in working with various emerging data sources, there are a substantial challenges as well. In the following sections, we outline some of the difficulties associated with accessing private data sources.

Sampling and Selection

The most significant challenge associated with using these data sources relates to working within an unknown sampling frame. Unlike with administrative data sets such as the Current Population Survey (CPS) or LEHD (Longitudinal Employer Household Data) data sets, the sampling frames associated with data sources generated by online labor intermediaries are often poorly understood. The different incentives for users to participate in online labor markets, knowledge exchanges, job boards and so on have implications for who appears in the sample, why and when they choose to contribute, the accuracy of the information that is provided, and therefore, the kinds of inference that can be drawn from the data generated by these websites. Kuhn and Skuterud (2004) report that workers who use job boards are positively selected on observables, but negatively selected on unobservables that might influence job search.[22] However, Kuhn and Mansour (2014), analyzing more recent data, find that these relationships may be changing.[24] They find that Internet job search lowers unemployment durations, suggesting that job seekers who use job boards are positively selected or that job search platforms have become better at being able to deliver jobs to unemployed workers.

Although characterization of the sampling frame can be difficult, it is sometimes possible to report at least basic comparisons of the sample statistics with those from data sets with sampling that is well understood. For instance, wages, education, age, gender, and so on can be compared with data sets such as the Current Population Survey, administered by the Census Bureau or with other supplementary data sources. Occupational distributions and geographic reach can be compared with sources such as the Occupational Employment Survey, administered by the Bureau of Labor Statics. This provides a basic understanding of how the sample might differ from an underlying population of interest.

Nevertheless, the severity of the sampling issues associated with most labor market intermediaries necessitates that researchers carefully match questions with the data sources generated by these websites in a way that mitigates issues related to the sampling frame.

Challenges in Arranging Access to the Data

Because these data are collected by private firms, they can be difficult to access. From the data provider's perspective, there are many costs associated with sharing data with researchers, including the potential for privacy intrusions which can hurt consumers or lead to negative public relations outcomes, as well as the time and cost required for the firm's technical employees to make the data available for analysis, either by training the researcher on the firm's systems or alternatively, by extracting the data in a way such that can be analyzed by researchers. Researchers, therefore, must make a compelling case to managers that the work that they would like to do is low cost to the data provider, and that it has potential benefits to the firm, for example in terms of informing business questions, generating positive effects for marketing or public relations, or for improving data quality.

In many cases, for data protection, firms prefer that researchers visit the firm and work onsite. This can be advantageous in terms of the scope, granularity of the data, and access to internal experts that is then available to researchers, but it can be costly for researchers who must be away from their home institutions. These costs of being off site are often higher than expected, because it limits access to colleagues and because productivity in other areas can be affected (e.g. progress on other papers and projects can slow down). There is also limited access to feedback on new ideas, as well as displacement costs and a number of unanticipated costs of not being around the researcher's home institution.

Finally, there is the potential for significant “red tape” at the researcher’s home institution, which in general, must to be navigated before any work begins. For instance, the analysis of corporate data often requires signatures from university officials on non-disclosure or legal agreements. Because these require the attention of different offices in other parts of the university or from higher level officers at the university, obtaining these approvals can be time consuming, and may require substantial lead time, normally counted in months. Moreover, the legal expectations of the data provider may differ substantially from what the university legal team is willing to accept, which can add time and risk to this process. For instance, data providers may have more stringent expectations in terms of who owns work product, whether they will have final review, how the data will be stored, and so on. The legal designation of the researcher--whether she is a contractor, employee, or something else--can also be a point of contention.

Finally, researchers also generally require Institutional Review Board (IRB) approval at their home institutions, which generates additional overhead especially in the context of experiments where subjects are “manipulated” online to understand a causal effect.

Challenges in Data Processing and Analysis

Although the data sources described above are disparate in their nature, most companies have their databases designed in ways that are broadly similar---tables of workers, employers, job openings, hours worked, contributions, and so on. Much of this data is stored in traditional off-the-shelf databases and accessing the data via SQL is straightforward. In this sense, these data sources are not "big" in the sense that special skills are required to extract or process the data.

However, most of the companies mentioned above also collect much more fine-grained data (such as user clickstream data, all actions that are taken within a mobile app, or GPS coordinates for every second a user is active on the platform). This ancillary data is often less structured and often is not in an analytically tractable format and may require special skills to access the data. Very often, it may be a database with an identifier, a timestamp and data stored in a lightweight interchange format (e.g. JSON). These data can be more difficult to work with and are often separated from the data that have direct business applications.

Due to the separation of business data across these formats as well as the size and scope of the data sets, there is significant time required for researchers to perform “data forensics”. Researchers must invest time in developing an understanding of where the data is high quality (i.e. where fields are well populated, data normalization, etc), and where the data are noisy or missing. Often, it is difficult to achieve an understanding of the data generating processes without access to the engineers responsible for designing the systems through which the data were collected. These factors can influence the types of questions that can be successfully answered using the data. Understanding the table structure, the primary keys, the foreign keys, and so on can take months before research projects can even begin.

This type of forensics, and the subsequent analysis, may also require an investment in technical skills. To effectively access and work with the data, it is often necessary to have at least a basic understanding of data manipulation languages such as Structured Query Language (SQL) or Apache Pig, Python, etc. to be able to access and manipulate the data. Due to time constraints on the firm's IT workers, there is generally little support that can be offered to researchers in terms of conducting the type of data manipulation necessary to get started with the data. Moreover, given the iterative nature of data forensics, outsourcing this to a technical assistant can be surprisingly difficult. Knowledge of the research question and the technical skills must often both reside within the same individual in order to make progress.

Publication Culture

An additional obstacle to establishing a successful research infrastructure through a corporate data partnership, especially when that partnership requires onsite work, is that the norms that academic scientists require for successful publication outcomes can be different than those that the company is able to support.

While scientists inside firms publish frequently, the existing cadre of data scientists is more likely to publish in Computer Science proceedings, which have much faster cycle times than social science publications that can take years in process. This raises the risk associated with social science projects because successful social science research outcomes, including the ability to respond to revision requests, can require access to a stable data set for several years. It is often difficult or impossible for high-tech firms to grant this type of access. One issue is that they may not permit researchers to create archives of the data, because it can violate the firm's privacy policy (i.e. we do not keep histories greater than 90 days). A second, and more common issue, is that it is very easy to lose access to the data when management or management policies change, or when key personnel leave, or when firms are acquired, disappear, and so on. Finally, if the data are available throughout the period of research collaboration, it is unlikely that firms are willing to make their data available for analysis in a way that is promoted by journals and the scientific community.

Conclusion

We are clearly entering a golden age for empirical labor market research. Data is becoming bigger and richer. There is a growing opportunity to revisit old questions with new and better data and to answer new questions raised directly by these new contexts. Therefore, even as technology continues to have a dramatic and disruptive effect on employment that commands greater policy and public attention, it can improve our ability to understand various market inefficiencies and to address policy concerns in a manner that is informed by rigorous analysis.

There are, however, substantial challenges to overcome. The data being collected by various labor intermediaries are scattered and heterogeneous and pose substantial practical, technical, and methodological challenges in order to be able to effectively use them for empirical research. However, if these challenges can be met, we expect these data sources to enable researchers to make substantial headway in the coming years into answering important questions about labor markets and employment that have historically been very difficult to approach using conventional data sources.

References

1. Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, *346*(6210), 1243089.

2. Autor, D. H. (2001). Wiring the labor market. *Journal of Economic Perspectives*, 25-40.

3. Marinescu, Ioana E. and Roland Rathelot (2015), “Mismatch Unemployment and the Geography of Job Search.”

4. Marinescu, Ioana E. and Ronald Wolthoff, (2015) “Opening the Black Box of the Matching Function: the Power of Words”

5. Tambe, P., & Hitt, L. M. (2013). Job hopping, information technology spillovers, and productivity growth. *Management science*, *60*(2), 338-355.

6. Ge, C., Huang, K. W., & Png, I. P. (2014). Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification Bias. *Online Profiles, and Misclassification Bias (November 27, 2014)*.

7. Tambe, P., & Hitt, L. M. (2012). The productivity of information technology investments: New evidence from IT labor data. *Information Systems Research*, *23*(3-part-1), 599-617.

8. Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science*, *60*(6), 1452-1469.

9. Fradkin, A. and Scott Baker (2015) The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data, Working Paper

10. Ghani, Ejaz, Kerr, William R. and Stanton, Christopher (2014) Diasporas and outsourcing: evidence from oDesk and India. Management Science, 60 (7). pp. 1677-1697. ISSN 0025-1909

11. Horton, John J. (2015) “Supply Constraints as a Matching Friction,” Working Paper

12. Horton, John J. (2015) “The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment” Working Paper

13. Horton, John J. and Ramesh Johari (2015) “At What Quality and What Price? Eliciting Buyer Preferences as a Market Design Problem,” Working Paper

14. Gee, L. (2015) “The More You Know: Information Effects in Job Application Rates by Gender In A Large Field Experiment”

15. Pallais A, Sands EG. Why the Referential Treatment? Evidence from Field Experiments on Referrals. Journal of Political Economy. Forthcoming.

16. Pallais, Amanda Inefficient Hiring in Entry-Level Labor Markets, American Economic Review. 2014;104(11):3565-3599.

17. Li, N. (2015). Labor Market Peer Firms. *Available at SSRN 2558271*.

18. Kokkodis Marios and Ipeirotis G. Panagiotis. (2015) “Reputation Transferability in Online Publications Labor Markets”. Management Science

19. Horton John J. (2015) “The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment” Working Paper

20. Bernstein, M., Little, G., Miller, R.C., Hartmann, B., Ackerman, M., Karger, D.R., Crowell, D., and Panovich, K. Soylent: A Word Processor with a Crowd Inside. In Proc. UIST 2010. ACM Press.

21. Stanton, Christopher and Thomas, Catherine (2014) *Landing the first job: the value of intermediaries in online hiring.* CEP Discussion Papers, CEPDP1316. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

22. Kuhn, P., & Skuterud, M. (2004). Internet Job Search and Unemployment Durations. *The American Economic Review*, *94*(1), 218.

23. Kuhn, P., & Mansour, H. (2014). Is Internet Job Search Still Ineffective?. *The Economic Journal*, *124*(581), 1213-1233.

Address correspondence to:

Prasanna Tambe
Information, Operations, and Management Sciences
Leonard N. Stern School of Business
New York University
44 West 4th Street, Rm 8-82
New York, NY 10012
Email: ptambe@stern.nyu.edu