Opening the Black Box of the Matching Function: the Power of Words *

Ioana Marinescu and Ronald Wolthoff †

Abstract

On the leading job board CareerBuilder.com, high-wage job postings unexpectedly attract *fewer* applicants, and this is the case even within a detailed occupation. Viewed through the lens of our directed search model, this negative relationship is indicative of substantial applicant heterogeneity within an occupation. Empirically, we find that job title heterogeneity is key: within a job title, jobs with 10% higher wages do attract 7.7% *more* applicants. Furthermore, our findings are consistent with a higher return to worker quality for hires in "manager" and "senior" job titles. Overall, our findings demonstrate the power of words in the matching process.

*We are grateful for comments from Briana Chang, Kerwin Charles, Francis Kramarz, Peter Kuhn, Alan Manning, Craig Riddell, and Chris Taber. This paper has benefited from feedback during presentations at CREST-Paris, the University of Chicago, the Annual Meeting of the Society of Economic Dynamics, the University of Toronto, the University of Wisconsin-Madison, UC Santa Barbara, the Tinbergen Institute, the University of British Columbia, the Bureau of Labor Statistics, the NBER Macro Perspectives Summer Institute. We thank Chaoran Chen and Sanggi Kim for excellent research assistance.

†Ioana Marinescu: School of Social Policy & Practice, University of Pennsylvania and NBER, ioana.marinescu@gmail.com; Ronald Wolthoff: Department of Economics, University of Toronto, ronald.p.wolthoff@gmail.com.

1 Introduction

A blossoming research agenda in macroeconomics analyzes business cycle fluctuations and public policy through the lens of the matching function (Petrongolo & Pissarides, 2001), which converts a given number of vacancies and unemployed workers into a resulting number of hires. This approach has been fruitful for improving our understanding of macro issues (Yashiv, 2007). Yet, there is limited empirical evidence on the micro foundations of the matching function, even though theory shows that these foundations are important for aggregate outcomes including unemployment and efficiency (Rogerson et al., 2005). Open questions include how employers advertise the nature of their jobs, and how workers react to this information[1]. Answering these questions requires opening the black box of the matching function.

Directed search models assign a key role to posted wages as an instrument for firms to attract the right pool of applicants (see e.g. Moen, 1997; Shimer, 2005; Eeckhout & Kircher, 2010). Yet, empirical evidence about employers' search strategies is limited. Holzer et al. (1991) show that the wages that firms pay affect the number of applicants that they attract, but the data in Holzer et al. (1991) does not capture whether firms communicated wages to potential applicants and the influence posted wages had on recruitment. In this paper, we investigate the role of posted wages in attracting applicants, and discuss how firm and worker heterogeneity can help us understand the search and matching process.

We investigate this question by using a directed search model and a new data set from CareerBuilder.com, a leading online job board. In our model, we focus on the role of firm and worker heterogeneity in explaining the relationship between posted wages and the number of applicants that a vacancy attracts. With only firm heterogeneity, high posted wages always attract more applicants: more productive firms post higher wages and attract more applicants. If we add worker heterogeneity, high posted wages can attract *fewer* applicants. We first discuss the case in which worker heterogeneity is horizontal, i.e. firms differ in their ranking of worker types. In this case, wages tend to be high and applicants tend to be few in labor markets where labor demand is high relative to labor supply (i.e. high tightness): in other terms, labor markets with high wages and a low number of applicants are those where workers are scarce. We then discuss the case in which worker heterogeneity is vertical, i.e. some workers are better in all jobs. In this case, high-wage jobs can attract fewer applicants if they are sensitive, i.e. if the productivity gap between high and low quality workers is higher than in low-wage jobs. An example of a sensitive type of job is a senior or manager position, while a less sensitive job is a junior position. The senior job title attracts a larger number of experienced applicants but

[1]See also Barron et al. (1985, 1987) about the screening and interviewing of applicants.

a much smaller number of inexperienced applicants, so that overall it attracts fewer but better applicants than a junior job title.

For our empirical analysis, we use data from CareerBuilder.com, which contains about a third of all US vacancies and is fairly representative of the US labor market. We use a data set of all the vacancies posted in Chicago and Washington, DC at the beginning of 2011. For each vacancy, we observe the information that firms provide in their job ads. We also have information on the pool of applicants that each job ad attracts, in particular the number of applicants and their education and work experience.

Using this data, we show that high-wage jobs attract significantly fewer applicants, and this negative relationship is robust to controlling for 6-digit SOC occupation fixed effects, and for firm fixed effects. At the same time, high-wage jobs attract more educated and more experienced applicants. Our model suggests that the negative relationship between wages and the number of applicants is driven by worker heterogeneity that is not captured by SOC fixed effects. We show that a job characteristic typically not considered in the economics literature plays a critical role in capturing worker heterogeneity. This piece of information is the *job title* of the vacant position as chosen by the employer, e.g. "senior accountant" or "network administrator". Within a job title, the relationship between wages and the number of applicants becomes positive instead of negative: a 10% increase in the posted wage is associated with a 7.7% increase in the number of applicants per 100 job views.

We analyze the reasons why job titles capture worker heterogeneity so much better than existing occupational classifications (SOC codes). We show that, relative to the detailed SOC occupations, job titles better reflect the hierarchy, level of experience, and specialization of different jobs. We find that the words in job titles associated with higher wages are also typically associated with fewer applicants. This contributes to explaining why, when job titles are not controlled for, we observe a negative relationship between wages and applicants within an SOC. Thus, our results uncover the previously undocumented power of words in the search and matching process.

Through the lens of our model, we can understand the role played by horizontal and vertical worker heterogeneity. Even within an occupation, there are important differences across jobs that can be explained by worker heterogeneity. As an example of horizontal heterogeneity within an SOC, outside sales jobs have higher posted wages and a lower number of applicants than inside sales. According to our model, this can reflect a higher demand for outside sales workers relative to supply. As an example of vertical heterogeneity within an SOC, manager jobs have higher posted wages and a lower number of applicants than junior jobs. This is consistent with manager jobs being more sensitive, in the sense that hiring a high-quality worker has higher returns for manager-type jobs than for junior-type jobs. The sensitivity of managerial jobs is consistent with prior literature showing that management can explain about 30% of differences in productivity across firms (Bloom et al., 2016).

Job titles not only play an important role in understanding worker application patterns, they also explain more than 90% of the variance in the midpoints of the wage ranges that firms post. By contrast, six-digit SOC codes, the most detailed occupational classification commonly used by economists, can only explain a third of this variance. The high explanatory power of job titles is not merely driven by the fact that there are more job titles than SOC codes: the adjusted R-squared is also close to 90% when controlling for job titles. Thus, employers advertise their jobs using the power of words embodied in the job title, and workers understand that jobs with different job titles are different. This suggests that the degree of frictional wage dispersion may be overestimated if one fails to control for job titles.

Our overarching conclusion is that words in job titles play a fundamental role in the initial stages of the search and matching process and are key to understanding labor market outcomes. We add to the literature in a number of ways.

First, our finding that job titles and posted wages affect the applicant pool that a firm attracts validates directed search models as realistic models of the labor market. Prior literature has found mixed evidence regarding the relationship between wages and the number of applicants.[2] We find a positive relationship, but only within a job title, which demonstrates that controlling for them is crucial. Our findings are consistent with the results in a number of recent studies, contemporaneous to our work, which also find a positive relation between wages and applicants of a similar magnitude (see Dal Bó et al., 2013; Banfi & Villena-Roldan, 2018; Belot et al., 2018).[3] We also document a positive relationship between wages and the quality of the applicant pool, consistent with evidence in Dal Bó et al. (2013).

Second, although job titles have been used to analyze career paths and promotions within firms (see e.g. Lazear, 1995), we are—to the best of our knowledge—the first to analyze their role in the search and matching process. Job titles can be seen as a new occupational classification which is based on employers' own description of their jobs rather than researchers' interpretation. We show that this new classification improves on existing occupation classifications (SOC) and has important implications for how we understand labor markets. We expect that this classification will prove to be a useful research tool—indeed, following our work, a

[2] Using a US survey, Holzer et al. (1991) document a non-monotonic relation, with jobs paying the minimum wage attracting more applicants than jobs that pay slightly less or slightly more. Using the full sample of the same data, Faberman & Menzio (2017) find a negative relation, even after controlling for occupation and industry.

[3] In recent years, various other approaches have been used to test whether search is directed, see e.g. Braun et al. (2016), Engelhardt & Rupert (2017) and Lentz et al. (2018). Somewhat more remotely, our results are also related to the literature on the elasticity of labor supply to the individual firm (see Manning, 2011, for a review).

number of recent papers have used job titles to obtain new insights about the labor market (e.g. Azar et al., 2017; Davis & Samaniego de la Parra, 2017; Banfi & Villena-Roldan, 2018).[4]

Finally, we add to the literature that analyzes wage variance. Most of this literature (e.g. Abowd et al., 1999; Woodcock, 2007; Abowd et al., 2002; Andrews et al., 2008; Iranzo et al., 2008; Woodcock, 2008) focuses on realized wages. It finds that unobserved characteristics captured by worker and firm fixed effects together explain most of the variance in realized wages (see e.g. Woodcock, 2007). We show that job titles explain as much of the variance in average posted wages as worker and firm fixed effects explain in realized wages. This suggests that *observable* job characteristics play an important role in explaining the wage variance.

This paper proceeds as follows. Section 2 presents our model and section 3 describes the data. Our main results are described in section 4. Section 5 provides additional results and robustness tests, after which section 6 concludes.

2 Model

In this section, we present a simple model of the labor market and discuss its predictions. The model classifies jobs in a hierarchy with two levels. Anticipating our empirical results, we name the broader classification *occupations* and the narrower classification *job titles*. Our model shows that a negative relation between wages and applications across job titles is indicative of worker heterogeneity and allows us to predict how horizontal vs. vertical worker heterogeneity can be captured in the data.

2.1 Setting

Segmentation by Occupation. We consider a static economy which is divided in disjoint segments and assume that only workers and firms in the same segment can produce output together. We equate a segment with an occupation, capturing the idea that a junior accountant may be able to do the job of a senior accountant, even if less well, but not the job of a nurse (and vice versa). The remainder of this section studies a particular occupation in isolation.

[4] Job titles seem particularly useful for the emerging literature on online job search (see e.g. Kuhn & Shen, 2012; Brenčič & Norris, 2012; Pallais, 2012; Faberman & Kudlyak, 2014; Gee, 2018; Marinescu, 2017). In addition, they may help to shed more light on topics like the gender and race wage gap (Blau, 1977; Groshen, 1991; Blau & Kahn, 2000), inter-industry wage differentials (Dickens & Katz, 1986; Krueger & Summers, 1986, 1988; Murphy & Topel, 1987; Gibbons & Katz, 1992), the specificity of human capital (Poletaev & Robinson, 2008; Kambourov & Manovskii, 2009), and occupational mobility and worker sorting (Groes et al., 2015).

Agents. Each occupation consists of a (normalized) measure 1 of firms, each with one vacancy, and a positive measure of workers, who each apply to one job.[5] Firms and workers are risk-neutral and maximize the product of their matching probability and their match payoff.

Workers differ in their skills and we distinguish between two types, indexed by $i \in \{0,1\}$. We denote the measure of workers of type i by μ_i.[6] Firms differ in two binary dimensions. The first dimension is represented by the job title $j \in \{A,B\}$, which codifies how a firm values the two types of workers, as we explain in more detail below. The second dimension is firms' productivity (or capital) $k \in \{L,H\}$, which we use to analyze heterogeneity in outcomes within a job title and which scales the output that a firm creates with any given worker. To simplify exposition, we assume that these characteristics are independent and equally common.[7] Hence, there is a measure $\frac{1}{4}$ of each of the four possible combinations of job title and productivity.

Search and Matching. Firms compete for workers by posting wages that are conditional on the worker's type (but not on their identity). That is, each firm of type (j,k) posts a menu of wages $\left(w_{0jk}, w_{1jk}\right)$.[8] After observing all wage menus, each worker submits one application to the firm that maximizes their expected payoff.[9] As standard in the literature, we assume that identical workers use symmetric strategies to capture the infeasibility of coordination in a large market. This assumption implies that the expected number of applicants of type i at a firm of type (j,k) follows a Poisson distribution with endogenous mean λ_{ijk}, which is known as the *queue length* (see e.g. Shimer, 2005).

Production and Payoffs. A match between a worker of type i and a firm of type (j,k) produces an output equal to the product of two components, $y_{ij}x_k$. The first component, $y_{ij} > 0$, represents the effect of the worker's skill and the firm's skill requirement, as codified by the firm's job title. We describe this component in more detail below. The second component of output, $x_k > 0$, captures the firm's productivity. Without loss of generality, we assume $x_H \geq x_L = 1$.

[5]We use 'firm', 'vacancy' and 'job' interchangeably. The same applies to 'worker' and 'applicant'.

[6]We treat workers' skills as exogenous. In a richer model with skill investment, skill heterogeneity can be supported through either stochastic returns to investment or heterogeneity in the costs of skill acquisition.

[7]These assumptions are not important for our results and can be relaxed by allowing for entry. In that case, firm heterogeneity can be supported by differences in entry costs or if firms learn their type after paying the entry cost.

[8]Since our empirical analysis focuses on firms that post wages, we restrict attention to these firms here. The possibility of applying to jobs without posted wages adds an outside option for workers, just like jobs outside CareerBuilder. For a model in which firms decide whether to post a wage or not, see Michelacci & Suarez (2006).

[9]The assumption of a single application per period is standard and captures the fact that there is an (opportunity) cost associated with applying. This cost prevents job seekers from applying to all desirable jobs, but instead forces them to carefully choose where to apply (see Belot et al., 2018). Galenianos & Kircher (2009) and Wolthoff (2018) develop models in which homogeneous workers send multiple applications. They find a positive relationship between wages and applications in equilibrium, which is consistent with our model's prediction within a job title.

Throughout, we assume that the difference between x_H and x_L is small enough to ensure that all firms receive applications. The worker's match payoff is their wage w_{ijk}, while the firm keeps the remaining output, i.e. $y_{ij}x_k - w_{ijk}$. Unmatched workers and firms get a zero payoff.

2.2 Empirical Predictions

We distinguish between three cases regarding match output y_{ij}. First, as a benchmark, we consider the case in which both types of workers are equally productive ("skill homogeneity"). Second, we discuss the case in which workers of type $i = 1$ are more productive in some jobs while workers of type $i = 0$ are more productive in other jobs ("horizontal differentiation"). Finally, we analyze the case in which workers of type $i = 1$ are more productive in all jobs ("vertical differentiation"). We summarize our results here and refer for additional details to online appendix A.

Skill Homogeneity. If the two types of workers are equally productive at any given firm, then $y_{0A} = y_{1A} \equiv y_A$ and $y_{0B} = y_{1B} \equiv y_B$, where we assume without loss of generality that $y_A \leq y_B$. That is, job titles are either noise ($y_A = y_B$), or represent a second form of firm heterogeneity ($y_A < y_B$) in addition to the heterogeneity in productivity x_k.

Within a job title j, we find that high-productivity firms (i.e. those with x_H) pay higher wages than low-productivity firms (x_L). That is, $w_{jH} > w_{jL}$. For workers to be indifferent between both types of firms, their matching probability must be lower at the high-productivity firms. In other words, high-productivity firms attract more applicants than low-productivity firms. Hence, the relation between wages and applications is positive within a job title j.

Across job titles, we distinguish two cases. Intuitively, if job titles are just noise (i.e., $y_B = y_A$), then both job titles offer the same wage and attract the same number of applicants. In contrast, if job titles represent firm heterogeneity, i.e. $y_B > y_A$, then the pattern is the same as within a job title: job title B offers higher wages and attracts more applicants than job title A. That is, the relation between wages and applications is positive across job titles.

Horizontal Differentiation of Skills. Suppose now that firms with job title $j = A$ rank workers in the opposite way of firms with job title $j = B$, because workers' types reflect skills that are specific to certain job titles. For example, a cardiology nurse is different from a neurology nurse. We formalize this with two assumptions: i) $y_{1B} \geq y_{0A} > 0$, which is without loss of generality, and ii) $y_{0B} = y_{1A} = 0$, which simplifies exposition because it implies that workers never apply to the job title in which they are less productive.[10]

It is straightforward to see that the empirical predictions within a job title are the same as with skill homogeneity: high-productivity firms attract more applicants and pay higher wages than low-productivity firms, making the relation between wages and applicants positive. Across job titles, the average number of applicants is given by $\frac{1}{2}\lambda_{0AH} + \frac{1}{2}\lambda_{0AL}$ at a firm with job title A, and an analogous expression holds for job title B. The ranking of this average number of applicants across job titles is ambiguous: job title B receives more applicants than job title A if there exist more workers of type 1 than of type 0, and vice versa. Intuitively, since the two job titles are equally common and workers perfectly sort themselves, the most-prevalent skill will generate the longest queue.

Unlike the number of applications, wages depend on the productivity of the worker-job pair y. As a result, wages could be higher or lower in the job title that attracts more applications. That is, the relationship between applications and wages could be positive or negative. Two forces are at play: labor market tightness and productivity. If the differences in productivity across job titles are small enough, i.e. $y_{0A} \to y_{1B}$, then the tightness factor dominates: firms offer higher wages in markets in which they expect few applicants, yielding a negative relationship between wages and applications. However, if productivity y is low enough in the market with fewer applications, it can overpower the tightness effect, producing an overall positive relationship between wages and applications.

For two horizontally differentiated job titles within an occupation, productivity differences may be small. In that case, one may expect a negative relationship between wages and applications across job titles. Ultimately, however, whether the tightness effect or the productivity effect dominates is an empirical question.

Vertical Differentiation of Skills. Finally, suppose all firms prefer one type of workers over the other type, e.g. because types reflect differences in experience or education. The analysis of this case resembles Faberman & Menzio (2017), but extends it with heterogeneity in firm productivity x. We use the terms *experienced* and *inexperienced* to distinguish between the two types of workers. Without loss of generality, we assume that the experienced workers are those with type $i = 1$. Hence, $y_{1j} \geq y_{0j}$ for $j \in \{A, B\}$. We also assume—again without loss of generality—that the difference in output between experienced and inexperienced workers

[10] The analysis remains similar if $0 < y_{0B} < y_{1B}$ and $0 < y_{1A} < y_{0A}$. The main difference is that some workers may start applying to the job title in which they are less productive if their output there is high or if the competition in 'own' job title is severe. However, the empirical predictions remain qualitatively unchanged. Note that the wage of the worker type that a firm does not attract in equilibrium is not uniquely pinned down: any wage that is sufficiently low will suffice. In practice, firms do not advertise wages for worker types that never apply, so we ignore those wages here.

is (weakly) larger for firms with job title $j = B$ than for firms with job title $j = A$, i.e. $\theta \equiv (y_{1B} - y_{0B}) / (y_{1A} - y_{0A}) \geq 1$. In line with Faberman & Menzio (2017), we will interpret θ as a measure of how sensitive job title B is relative to job title A: the higher θ, the more job title B gains from hiring an experienced worker instead of an inexperienced worker, compared to job title A. In general, we expect more senior jobs to be more sensitive. We focus on values of θ that guarantee that both job titles attract both types of workers.

Within a job title, we obtain the same results as before: high-productivity firms pay higher wages and attract more applicants than low-productivity firms, making the relation between wages and applications positive. However, because heterogeneity is vertical, the model now yields an additional prediction, which concerns the quality of the applicant pool. In particular, we find that within a job title firms that post higher wages attract more experienced applicants.

Across job titles, the relationship regarding the number of applicants can be negative. In the appendix, we show that this is the case if job title B is i) less productive with an inexperienced worker than job title A and ii) sufficiently sensitive. In that case, job title B pays higher wages and attracts a smaller, but better, more experienced pool of applicants.

Summary. The model implies that observing a negative relationship between wages and applications across job titles within an occupation is indicative of the presence of heterogeneity in worker skill on top of the heterogeneity in firm productivity. This worker heterogeneity can be either horizontal or vertical. What we learn differs between these two cases. With horizontal differentiation, a negative relationship indicates that productivity differences between job titles are small and that wage differences arise primarily due to tightness: firms in a job title with fewer applications face more competition and therefore must pay higher wages. With vertical differentiation, the negative relationship between wages and applications is informative about job sensitivity: job titles with higher wages but fewer applicants can be identified as sensitive job titles, where the benefits of hiring more experienced workers are greater.

3 Data

To analyze the relation between wages and the pool of applicants within and across job titles, we use proprietary data provided by CareerBuilder.com. In this section, we describe the main features of our data set.

Background. CareerBuilder is the largest online job board in the United States, visited by approximately 11 million unique job seekers during January 2011.[11] While job seekers can use the site for free, CareerBuilder charges firms several hundred dollars to post a job ad on the website for one position for one month. A firm that wishes to keep the ad online for another month is subject to the same fee, while a firm that wishes to advertise multiple positions needs to pay for each position separately, although small quantity discounts are available (see CareerBuilder, 2013). At each moment in time, the CareerBuilder website contains over 1 million jobs.[12]

Search Process. A firm posting a job is asked to provide various pieces of information. First, it needs to specify a job title, e.g. "senior accountant," which will appear at the top of the job posting as well as in the search results. CareerBuilder encourages the firm to use simple, recognizable job titles and avoid abbreviations, but firms are free to choose any job title they desire. Further, the firm provides the full text of the job ad, a job category and industry, and the geographical location of the position. Finally, the firm can specify education and experience requirements as well as the salary that it is willing to pay.[13]

Job seekers who visit CareerBuilder.com see a web form which allows them to specify some keywords (typically the job title), a location, and a category (broad type of job selected from a drop-down menu). After providing this information[14], job seekers are presented a list of vacancies matching their query, organized into 25 results per page. CareerBuilder sorts the job ads by an index called 'relevance', which is determined by a proprietary algorithm that aims to describe the fit between the position and the job seeker. Job seekers can change the default sort order and sort the ads by company, distance or posting date instead. For the jobs that appear in the list with search results, the job seeker can see the job title, salary, location, and the name of the firm. To get more details about a job, the worker must click on the job snippet in the list, which brings them to a page with the full text of the job ad as well as a "job snapshot" summarizing the job's key characteristics. At the top and bottom of each job ad, a large "Apply Now" button is present, which brings the worker to a page where they can send their resume and their cover letter to the employer.

[11] See comScore Media Metrix (2011). Monster.com is similar in size, and whether Monster or CareerBuilder is larger depends on the exact measure used.

[12] Our data is limited to search activity on CareerBuilder. We therefore miss information on search activities on other employment websites or offline. See e.g. Nakamura et al. (2008), Stevenson (2008) and Kuhn & Mansour (2014) for studies of search behavior across platforms.

[13] CareerBuilder provides no guidance in these choices, although firms can of course observe other ads on the site.

[14] It is not necessary to provide information in all three fields.

Sample. Our data set consists of vacancies posted on CareerBuilder in the Chicago and Washington, DC Designated Market Areas (DMA) between January and March 2011. A DMA is a geographical region set up by the A.C. Nielsen Company that consists of all the counties that make up a city's television viewing area. DMAs are slightly larger in size than Metropolitan Statistical Areas and they include rural zones. Our data is a flow sample: we observe all vacancies posted in these two locations during January and February 2011 and we observe a random subsample of the vacancies posted in March 2011.

Variables. The CareerBuilder data is an attractive source of information compared to existing data sets, in particular due to the large number of variables that it includes. For each vacancy, we observe the following *job characteristics*: the job title, the salary (if specified), whether the salary is hourly or annual, the education level required for the position, the experience level required for the position, an occupation code, and the number of days the vacancy has been posted. The occupation code is the detailed, six-digit O*NET-SOC (Standard Occupational Classification) code.[15] CareerBuilder assigns this code based on the full content of the job ad using O*NET-SOC AutoCoder, the publicly available tool endorsed by the Bureau of Labor Statistics.[16] This procedure is consistent with the approach of the Current Population Survey (CPS).[17] We further observe the following *firm characteristics*: the name of the firm, an industry code, and the total number of employees in the firm. CareerBuilder uses external data sets, such as Dun & Bradstreet, to match the two-digit NAICS (North American Industry Classification System) industry code and the number of employees of the firm to the data.

In addition to these characteristics, we also observe several *outcome variables* for each vacancy. Our first outcome variable, the number of views, represents the number of times that a job appeared in a listing after a search. The second outcome variable, the number of clicks, is the number of times that a job seeker clicked on the snippet to see the entire job ad. Finally, we observe the number of applications to each job, where an application is defined as a person clicking on the "Apply Now" button in the job ad.

From these numbers, we construct two new variables that reflect applicant behavior: the number of applications per 100 views, and the number of clicks per 100 views. These measures correct for heterogeneity in the number of times a job appears in a listing, allowing us to analyze applicants' choices among known options.

[15]See `http://www.onetcenter.org/taxonomy.html`. We henceforth refer to this classification simply as SOC.

[16]See `http://www.onetsocautocoder.com/plus/onetmatch`.

[17]This means that misclassification is unlikely to be a larger problem in the CareerBuilder data than in the CPS. See Mellow & Sider (1983) for an analysis of inconsistencies in occupational codes in the CPS.

For a random subset of the vacancies of January and March 2011, we also observe *applicant characteristics*. Specifically, we observe the number of applicants broken down by education level (if at least an associate degree) and by general work experience (in bins of 5 years). We will use these job seeker characteristics to analyze the quality of the applicant pool that a firm attracts.

Cleaning. We express all salaries in yearly amounts, assuming a full-time work schedule. When a salary range is provided, we generally use the midpoint in our analysis, but we perform robustness checks in appendix C. For ease of exposition, we will refer to this midpoint as the *wage* of the job. To reduce the impact of outliers and errors, we clean the wage data by removing the bottom and top 0.5%.

Because job titles are free-form, many unique ones exist and the frequency distribution is highly skewed to the right. To improve consistency, we cleaned the data. Most importantly, we formatted every title in lower case and removed any punctuation signs, employer names, or job locations. In most of our analysis, we restrict attention to the first four words of a job title. As we will discuss, because this restriction has minimal impact on the number of unique job titles in our sample, our results are not sensitive to it.

Representativeness. Some background work (data not shown) was done to compare the job vacancies on CareerBuilder.com with data on job vacancies in the representative JOLTS (Job Openings and Labor Turnover Survey). The number of vacancies on CareerBuilder.com represents 35% of the total number of vacancies in the US in January 2011 as counted in JOLTS. Compared to the distribution of vacancies across industries in JOLTS, some industries are overrepresented in the CareerBuilder data, in particular information technology; finance and insurance; and real estate, rental, and leasing. The most underrepresented industries are state and local government, accommodation and food services, other services, and construction.

While CareerBuilder data is not representative by industry, in most other respects it is representative of the US labor market. Using a representative sample of vacancies and job seekers from CareerBuilder.com in 2012, Marinescu & Rathelot (2018) show that the distribution of vacancies across occupations is essentially identical (correlation of 0.95) to the distribution of vacancies across all jobs on the Internet as captured by the Help Wanted Online data. Furthermore, the distribution of unemployed job seekers on CareerBuilder.com across occupations is similar to that of the nationally representative Current Population Survey (correlation of more than 0.7). Hence, the vacancies and job seekers on CareerBuilder.com are broadly representative of the US economy as a whole, and they form a substantial fraction of the market.

Descriptive Statistics. Table 1 shows the summary statistics for our sample. The full sample consists of more than 60,000 job openings by 4,797 different firms. On average, each job was online for 16 days, during which it was viewed as a part of a search result 6,084 times, received 281 clicks, and garnered 59 applications. Per 100 views, the average job receives almost six clicks and approximately one application.[18]

Only a minority of job ads include an explicit experience requirement (0.3%) or an explicit education requirement (42%). When specified, these requirements appear in the "job snapshot" box at the end of the full job ad, but they do not appear in the job snippet that job seekers first see in the search results. Therefore, employers may choose not to fill in education and experience requirements if they feel that the overall job description is sufficiently informative.[19]

We observe wage information for approximately 20% of the jobs, from 1,369 unique firms. When present, the wage information appears in the job snippet as part of the search results. In 87% of these cases, a salary range is provided, with a width of 25% of the midpoint on average. When posting a range, firms generally state that pay is "commensurate with candidate qualifications and experience", so not all wages in the range are equally likely for a given job seeker. The average posted yearly salary is just over $57,000, and we will show in more detail below that posted wages on the website have the same distribution as the wages of full-time workers in the Current Population Survey. Finally, we observe the average applicant quality for approximately 2,300 job . The average applicant has between 16 and 17 years of education (conditional on holding at least an associate degree) and just over 13 years of work experience.

Job Titles. All job ads in our sample specify a job title. This job title is prominently featured on the employment website and is the main piece of information that workers use to search the CareerBuilder database. The full sample contains 22,009 unique job titles. Truncation to the first four words marginally reduces this number to 20,447. In the subsample of jobs with posted wages, the corresponding numbers are 4,669 and 4,553, respectively. In Table 2, we list the ten most common job titles (after truncation), both for the full sample and for the subsample of jobs that post wages. Note that the most common job titles are typically at most three words long. We also show the most common job titles if we truncate the job title to the first two words or the first word. Figure 1 provides a more comprehensive overview of frequent job titles in the form of a word cloud, in which the size of a job title depends on its frequency.

Inspection of the table and the figures reveals that job titles often describe occupations, e.g. "administrative assistant," "customer service representative," or "senior accountant." This

[18]Keep in mind that the average of ratios does not necessarily need to equal the ratio of averages.

[19]In an alternative data set, Hershbein & Kahn (2016) find that as much as half of all firms post an education and/or experience requirement.

raises the question of how job titles compare to other definitions of occupations, in particular the occupational classification (SOC) of the Bureau of Labor Statistics. Perhaps the most obvious difference between job titles and SOC codes concerns their variety: in our full sample, the number of unique job titles is more than 25 times the number of unique SOC codes.[20] In other words, job titles provide a finer classification. For example, they distinguish between "inside sales representative" and "outside sales representative," between "executive assistant" and "administrative assistant," and between "senior accountant" and "staff accoutant"—while each of the two jobs in these pairs has the same SOC code. While some of the larger variety in job titles might be due to noise in employers' word choice, we will show in the following sections that distinctions between job titles are economically significant.

4 Empirical Analysis

Our empirical analysis consists of two parts. First, we analyze how the wage that a firm posts affects its number of applicants. Subsequently, we analyze the effect of the wage on the average quality of applicants. We interpret the results through the lens of our model.

4.1 Number of Applicants

Table 3 presents our first set of results. As we discuss in more detail below, the set of controls varies across the columns, with log wages always being included. The dependent variable in each specification is the number of applications per 100 views, which we use to correct for heterogeneity in the number of views across jobs.[21]

4.1.1 Wage Impact

Across Job Titles. We start by exploring the relation between the wage and the number of applicants without any additional controls (column I). This cross-sectional relationship is significantly negative in our sample: a 10% increase in the wage is associated with a 6.3% decline in applications per view.[22] As highlighted by our model, this negative relation is indicative of worker heterogeneity: after all, it is perfectly possible that a firm looking for an accountant

[20] 20,447 versus 762, to be exact. In the subsample with posted wages, the difference is smaller, but still a factor of eight (4,553 versus 594). Note that the SOC classification distinguishes 840 occupations in total, some of which do not appear in our data.

[21] An alternative choice for the outcome variable is simply the logarithm of the number of applicants for each job. We find that our key results from Table 3 are qualitatively unaffected by this alternative outcome definition.

[22] A 10% increase in the wage decreases the number of applications per 100 views by $0.770 \log(1.1) = 0.073$, which is a 6.3% decline compared to the sample average of 1.168.

offers a higher wage but attracts fewer applicants than a firm looking for a customer service representative.

More interestingly, however, we find that the negative association between wages and the number of applicants survives if we add controls that are commonly used to capture labor market heterogeneity, such as job characteristics as well as detailed occupation fixed effects (column II) and firm fixed effects (column III). These controls indeed explain part of the variance in the number of applicants, as demonstrated by the increase in the R^2, but the coefficient on the wage and its significance remain essentially unchanged: in column III, an 10% increase in the wage offer is associated with a statistically significant 5.8% decline in the number of applications per view. In light of our model, the fact that the coefficient remains negative indicates that there is meaningful skill heterogeneity even within detailed occupations and firms.

Within Job Titles. As job titles provide a finer classification of jobs than SOC codes, they should better capture skill heterogeneity. If job titles capture worker heterogeneity sufficiently well, we should see a positive relationship between wages and applications within job title, as predicted by our model.

We test this hypothesis in column IV and V by replacing the SOC code fixed effects with job title fixed effects, allowing us to consider the relation between wages and applications within a job title. This exercise yields fundamentally different results. In particular, the negative relationship between wages and the number of applications seen in columns I, II, and III now becomes significantly positive. That is, within job title, higher wages are associated with more applicants. This is true regardless of whether we include firm fixed effects or not. The point estimate in column V implies that a 10% increase in the wage is associated with a 7.7% increase in the number of applicants per 100 views.

As the R^2 indicates, the specifications with job title fixed effects explain a larger part of the variation in the number of applicants per view. At some level, this is not surprising as there are many more job titles than occupations. However, that is not the full story since measures that correct for the larger number of controls, such as the adjusted R^2 and the AIC, also favor the specifications with job titles. The combination of these results strongly suggests that job titles capture meaningful worker heterogeneity that is glossed over by standard occupational codes.

4.1.2 Word Analysis

While it is informative to know that controlling for job title fixed effects reverses the sign of the relationship between wages and the number of applicants, this fact in itself does not reveal what heterogeneity is captured by job titles: as our model showed, both vertical and horizontal worker heterogeneity are consistent with a negative relationship between wages and the number of applicants. To better understand the nature of the heterogeneity, we analyze what information is contained in the job titles. We do so by using fixed effects for each separate word.

Specifically, we regress wages and the number of applicants per view on detailed SOC codes, compute the residuals, and regress those residuals on a set of dummy variables for each word appearing in the job title. Compared to job title fixed effects, this specification is restrictive because it ignores the order and combinations in which words appear in a job title; it assumes, for example, that the word “assistant” has the same effect in “executive assistant” and “assistant store manager”. Yet, this specification allows us to determine which words are most important.

Wages. First, we explore which words are most important in explaining wage variation within SOCs. In Table 4, we list words that appear at least 100 times and are significant at least at the 5% level. We checked the job titles in which these words appear and manually classified the words into three categories.

The first column includes words that suggest the presence of vertically differentiated skills within an SOC code, as these words signal the seniority of the worker holding this job title. Within an SOC code, job titles that include the words “manager” or “senior” have higher than average posted wages, whereas wages are lower than average for titles that include the words “specialist” or “junior”. For instance, within the SOC code 13-2011 (“Accountants and Auditors”), accounting managers and senior accountants earn more than accounting specialists and junior accountants.

In the second column, we list words that suggest the presence of horizontally differentiated skills within an SOC, as these words capture specialties or skills. For example, within the SOC code 41-3099 (“Sales Representatives, Services, All Other”) or 41-4012 (“Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products”), inside sales jobs, which require employees to contact customers by phone, pay less than outside sales jobs, where employees must travel and meet face-to-face with customers.[23] Finally, the third column is similar to the second column, but focuses on computer skills and specialties. For example, within SOC code 15-1071 (“Network and Computer Systems Administrators”), network administrators earn less than systems administrators.

Figure B.1 in the online appendix provides a more complete overview of the words that

[23]The most frequent word among those that are associated with higher wages in the second column is "-". This is not a typo; this character typically separates the main job title from additional details about the job. These additional details were deemed important enough for the firm to specify them in the job title. Presumably, all other things equal, a more specialized job comes with a higher pay. Some examples of the use of "-" are: "web developer - c# developer - net developer - vb net developer" or "web developer - ruby developer - php developer - ror pearl java".

are associated with either higher or lower wages. The size of the words represents their frequency, while the intensity of the color represents the magnitude of their effect. We can classify the words in this figure in roughly the same way as the frequent words from Table 4. For example, "president" and "intern" indicate very different levels of seniority and have opposing effects on the wage within an SOC, just as one would expect. "Scientist" and "retail" are examples of skills/specialties leading to a higher and a lower wage, respectively. Finally, in terms of computer skills, the word "linux" is associated with higher wages, while the generic term "computer" leads to lower wages within an SOC.

Applications. We now discuss words that are associated with a larger or smaller number of applicants per view within SOC (Figure B.2 in the online appendix). Many of the words that predict a greater number of applicants per view are also words that predict lower wages (compare with Figure B.1). These include words denoting lower seniority and experience (vertical differentiation), such as "assistant", "support", "specialist", "coordinator", "entry", and "junior". As for words denoting specialties (horizontal differentiation), we can see, for example, that lower-wage inside sales jobs receive more applications than the average job in the same SOC.

Conversely, many of the words that predict a lower number of applicants per view within SOC are words that also predict higher wages. The two word clouds are remarkably similar. Words that denote higher seniority and management positions such as "manager", "senior", and "director" are associated with a smaller number of applicants. Words that are associated with higher paying specialties or areas, such as "developer", "engineer", and "linux", have a lower number of applicants.

Overall, examining the words that predict wages and words that predict the number of applicants enlightens us on the negative relationship between wages and applicants within SOC. Within SOC, words in the job title associated with higher wages predict a lower number of applicants per view, and vice versa for words in the job title associated with lower wages. Substantively, jobs with higher seniority or managerial responsibilities (vertical differentiation) tend to pay higher wages and attract fewer applicants. Similarly, specialties (horizontal differentiation) with higher pay tend to attract fewer applicants.

4.1.3 Interpretation

The empirical results indicate that our model's classification of jobs into a hierarchy of job titles within occupations is economically meaningful. Given the detailed nature of six-digit SOC codes, a natural conjecture would have been that they capture worker heterogeneity quite well. In our model, however, this would have implied a positive relation between wages and applications within an SOC. The fact that this relation is negative within detailed SOC codes implies that there remains substantial worker heterogeneity at this level.

Our model predicts that the relation between wages and applications should become positive once we control for this heterogeneity. The fact that this happens when we control for job title fixed effects indicates that job titles capture an important part of the worker heterogeneity. Furthermore, the words in the job titles provide insight into the nature of the heterogeneity.

Our model provides a clear way to interpret a negative relation between wages and applications across job titles. As discussed in section 2, a negative wage-application relation between two horizontally differentiated job titles indicates that the relative difference in market tightness dominates the relative difference in productivity. This suggests, for example, that outside sales representatives earn more than inside sales representatives not because they are substantially more productive, but because there is more demand for outsides sales representatives relative to supply.[24]

Further, a negative wage-application relation between two vertically differentiated job titles indicates that the jobs with the higher wages are sensitive in the sense that they benefit disproportionally from hiring an experienced worker rather than an inexperienced worker. Through this lens, we find that senior or executive jobs (e.g. "senior accountant") are examples of sensitive jobs, while assistant, associate, junior or coordinator jobs (e.g. "junior acountant") are examples of more regular jobs. We also find manager jobs to be sensitive, which is consistent with the literature on the impact of management on firm performance. For example, Bloom et al. (2016) show that differences in management practices account for about 30% of total factor productivity differences across firms. Bender et al. (2017) find that better-managed firms systematically recruit and retain better workers, which is an important determinant of these productivity differences. This leveraging impact on firm productivity through workforce selection suggests that hiring a better manager over a worse one is more important than hiring a better "junior" person over a worse one, which is precisely our notion of sensitivity.

4.2 Quality of Applicants

As discussed in section 2, our model yields predictions not only about the number of applicants that a firm attracts, but also about their types. Our data contains two pieces of information about applicants' characteristics: i) their work experience, and ii) their education level, expressed in

[24] Davis & Marinescu (2018) provide more direct evidence for the tightness effect by showing that labor market tightness measured as vacancies over applications has a positive impact on posted wages in data from CareerBuilder.com.

years of education. Interpreting these variables as proxies of vertically differentiated skills, we analyze how they vary with the wages that firms post.

4.2.1 Wage Impact

Average Experience. The first two columns in Table 5 displays the results for the average experience level. Not surprisingly, we find that in the cross-section higher wages are associated with more experienced applicants (column I). More interestingly, the relation survives once we control for job title fixed effects and job characteristics (column II) although both the magnitude and the significance level are somewhat reduced.[25] To be precise, the specification with job title fixed effects and job characteristics indicates that a 10% increase in the wage is associated with an increase in the experience of the average applicant by 0.15 years, or roughly 1%.

Average Education. In column III and IV of Table 5, we focus on the education level of the average applicant.[26] The results are very similar to what we found when explaining the experience of the average applicant. First, higher wages are associated with more educated applicants in the cross-section (column III). Second, after controlling for job title fixed effects and job characteristics, this effect remains although the magnitude and the significance are somewhat reduced (column IV). Quantitatively, the effect is small with a 10% increase in the wage being associated with an increase in the number of years of education of less than 1% within a job title.

4.2.2 Word Analysis

As before, we develop a better understanding for the importance of job titles by investigating which words are particularly important for explaining the variation in the quality of applicants within an SOC code. We report the results in Table B.1 in the online appendix.

Words that predict higher experience or higher education tend to be words that also predict higher wages, and vice versa for words that predict lower experience or education. For example, words that indicate higher seniority or management such as "manager" and "senior" are associated with higher experience, while words like "director" and "chief" are associated with both higher experience and higher education. Lower education and experience are associated with certain specialties. The example of "rn" (registered nurse) is interesting, as it is associated

[25]One factor that may affect the significance of our estimates is the relatively small sample size and the large number of job titles.

[26]See also Kudlyak et al. (2012) for evidence on how different jobs attract workers with different levels of education.

with both lower experience and lower education. This is explained by the fact that, within SOC 29-1111 ("Registered Nurses"), "rn" indicates a lower level job compared to job titles where "rn" does not appear, such as "nurse manager," "nurse clinician," and "director of nursing."

4.2.3 Interpretation

Our word analysis shows that words predicting more experience or education tend to be words that also predict higher wages. This explains why the effect of the wage on experience and education in Table 5 is reduced once we control for job title fixed effects.

These results are in line with our model for jobs with vertically differentiated skill requirements. Higher education or experience corresponds to the notion of more "experienced" workers in the model: such workers are preferred by all firms. When sensitive jobs (e.g. "senior accountant") pay higher wages but receive fewer applicants than non-sensitive jobs (e.g. "junior accountant"), our model predicts that the sensitive job should attract an applicant pool that is more "experienced" on average. This is exactly what we observe for words like "senior", "manager", "director", "executive", "management", et cetera. The fact that a positive correlation between wages and the quality of applicants exists *within* a job title is also consistent with our model with vertically differentiated skills: more productive firms within a job title pay higher wages and attract more experienced workers.

5 Additional Results and Robustness

5.1 Number of Clicks

We have analyzed the impact of wages on the number of applicants and have shown that this impact is only estimated to be positive when controlling for job titles. However, omitted variable bias could contaminate the relationship between wages and the number of applications: since we cannot control for the full text of the job ad, we may be missing information that is relevant for the worker's application decision. To assess whether this is the case, we turn to an examination of the impact of the wage on the the number of times potential applicants *click* on a job ad for more info (per 100 views). Recall that when a job is listed as a snippet on the result page, only the posted wage, job title, firm, and DMA are listed. The applicant must click to see more details. Hence, we have all the variables that can drive the applicant's click decision, eliminating the scope for omitted variable bias.

Table 6 explores the relationship between wages and clicks per 100 views: the results are similar to those obtained when applications per 100 views is the dependent variable (Table 3).

When no controls are used (column I), we see a significant and negative association between the wage and clicks per 100 views. When controlling for basic job characteristics, firm fixed effects, and job title fixed effects, the coefficient on the wage becomes positive and highly significant (column V), implying that a 10% increase in the wage is associated with a 2.9% increase in clicks per 100 views. The fact that the qualitative results in Table 3 can be reproduced for clicks per view, an outcome whose determinants are fully known, improves our confidence in our basic results.

5.2 Wage Posting

Our results so far are based on the 20% subset of job ads containing an explicit announcement regarding the wage, which is the relevant sample for trying to understand whether workers direct their search to higher wages.[27] However, a natural question is how this sample compares to jobs on CareerBuilder without a wage.

We analyze this question in online appendix C.1. We find that job titles and firm fixed effects each explain around 70% of the variance in wage-posting behavior, and that together they explain essentially all of the variation (93%). Job characteristics only have a small impact on the likelihood that a firm posts a wage, although the effect is sometimes statistically significant (e.g. in the case of educational requirements). We further analyze which words significantly increase or decrease the probability that a job ad contains a wage. These words do not appear to be systematically related to the wages that firms pay in the sense that both "high-wage" words and "low-wage" words from Table 4 can predict a higher probability of posting a wage.

5.3 Wage Offers

A further natural question is how the distribution of posted wages on CareerBuilder compares to the cross-sectional distribution of realized wages in the US. To answer this question, we use data from the basic monthly CPS from January and February 2011. We restrict the CPS data to employed individuals in the Chicago and Washington, DC MSAs, such that the sample covers

[27] Various explanations for the fact that not all firms post a wage are possible. Some firms may not wish to commit to a particular wage ex ante or may feel that posting a wage is unnecessary because the rest of the job ad provides sufficiently precise information on compensation. Alternatively, some companies may use Applicant Tracking Systems (ATS) software that keeps track of job postings and application, but typically removes the wage information by default before sending the job posting to CareerBuilder (private communication with CareerBuilder.com). However, this is unlikely the full story, as the fact that most job ads do not advertise wages is consistent with the worker survey data of Hall & Krueger (2012) and evidence from job boards in other countries where ATS may be less common (Brencic, 2012; Kuhn & Shen, 2012). For more discussion of wage posting on online platforms, see also Brown & Matsa (2016).

approximately the same time frame and geographic area as the CareerBuilder data.

Cross-Sectional Distribution. The results of the comparison between the two data sets are displayed in Figure 2. The upper panel shows that the posted wage distribution is more compressed than the realized wage distribution. However, the CareerBuilder data does not properly distinguish full-time from part-time jobs: in particular, hourly wages are converted to full-time equivalent. Furthermore, CareerBuilder data does not account for sporadic patterns of employment that could occur for some workers in the CPS. Therefore, posted wages mostly capture full-time work. Another difference between the CPS and CareerBuilder is that CPS earnings are top-coded. To make the two data sets more comparable, we restrict the CPS data to workers who work full-time and whose earnings are not top-coded, and the CareerBuilder data to earnings levels that are not top-coded in the CPS. The lower panel in Figure 2 shows the resulting distributions. We find that—despite the fact that posted wages are not always observed or accepted and can sometimes be renegotiated[28]—the distribution of posted wages on CareerBuilder is now nearly identical to the distribution of realized wages in the CPS.[29]

Effect of Occupations. To further compare the CareerBuilder data to the CPS, we investigate the effect of occupational controls in both data sets. We present the detailed results of this exercise in appendix C.2 and briefly summarize them here.

For the CPS, we regress log weekly earnings on increasingly finer occupation controls. We find that the most aggregated classification (*major* occupations), distinguishing 11 different occupations, explains approximately 15% of the variation in the wages. When we use specifications with 23 *minor* and then 523 *detailed* occupations, the adjusted R^2 rises to 18% and 36%, respectively, still leaving most of the wage variance unexplained.

We repeat this exercise for the CareerBuilder data to analyze the degree to which SOC codes can explain the variance in *posted* wages, i.e. the midpoint of the posted salary ranges. We find that the results are strikingly similar to the CPS sample: the adjusted R^2 is 14% for major occupations, 17% for minor occupations and 39% for detailed occupations. This similarity between the explanatory power of occupations further supports the idea that the posted wages in our data are roughly comparable to the realized wages in the CPS.

The CareerBuilder data of course allows us to go further and control for job titles. We find that this exercise explains more than 90% of the variance in posted wages, meaning that

[28]See e.g. Andrews et al. (2001) for evidence on the incidence of renegotiation.

[29]Of course, the two distributions do not need to coincide. It is straightforward to specify a search model where realized wages first-order stochastically dominate posted wages. The goal here is merely to show that posted wages at CareerBuilder are similar to wages in the labor market as a whole.

relatively little variation in posted wages is left within a job title[30]. Since the explanatory power of occupations is essentially the same in the CPS and CareerBuilder samples, it is possible that job titles, were they available in the CPS, would explain much of the variance in realized wages. These results suggest that existing estimates of the degree of frictional wage dispersion (see e.g. Hornstein et al., 2007) may be too high. Revisiting these estimates requires knowledge of the extent to which a jobs with different job titles are substitutable for a given worker. This is an exciting area for future research.

5.4 Occupation Weights

Finally, we investigate the robustness of our results to reweighing our data to make it representative of the universe of US jobs advertised online. We construct pseudo-sampling weights to correct for any discrepancy between the distribution of SOC codes in our data and the corresponding distribution among online vacancies more broadly. Our source for the latter distribution is the The Conference Board Help Wanted Online (HWOL) Data Series for January-February 2011. For each SOC, the pseudo-sampling weight is defined as the ratio of the number of vacancies in that occupation in our data to the total number of vacancies in that occupation in the HWOL.

Table 7 presents the results of our main specification using these weights. We find that our main result is unchanged: a higher wage is associated with fewer applications per search within SOC, but more applications per search within job title (compare to Table 3). Furthermore, the magnitude of the coefficient on the posted wage is very close to what is shown in Table 3.

6 Conclusion

In this paper, we start with developing a theoretical model that examines the role of firm and worker heterogeneity in the matching process. If only firms are heterogeneous, high wage jobs at more productive firms receive more applications. However, when workers are also heterogeneous, high wage jobs can receive fewer applications than low wage jobs. With horizontal heterogeneity, higher labor market tightness in the market for a specific worker type is associated with both higher wages and fewer applicants. With vertical heterogeneity, higher wage

[30] A natural concern may be that the large explanatory power of job titles is partially mechanical as there are many job titles in our data. We explore this hypothesis in appendix C.2 in a number of ways: we perform a permutation test, we limit the sample to frequent job titles, we explore specifications with fewer fixed effects, and we explore alternative definitions of the wage. All our results there indicate that the mechanical part of the effect is small.

jobs can attract fewer applicants if they are more sensitive, i.e. if the productivity gap between high and low quality workers is higher.

We have used new data from CareerBuilder.com to show that within an SOC code, higher paying jobs attract *fewer* applicants; the relationship between posted wages and the number of applicants only becomes positive within job titles. Specifically, within a job title, a 10% increase in posted wages is associated with a 7.7% increase in applications per 100 job views. Using the insights from our model, this implies that job titles capture important worker heterogeneity missed by 6-digit SOC codes. Using word analysis, we determine that this heterogeneity can be either horizontal, as in the case of inside sales vs. outside sales, or vertical as in the case of junior accountants vs. senior accountants. Furthermore, our model and data taken together imply that, within an occupation, jobs that include managerial duties are sensitive, i.e. there are substantive productivity returns to hiring high-quality workers over low-quality workers.

Another way of understanding the importance of job titles in capturing heterogeneity is to regress posted wages and the quality of applicants on job titles. Job titles explain 90% of the variance in posted wages, while detailed SOC codes only explain about a third of the variance. Job titles also explain more than 90% of the variance in the average education and experience of applicants that a vacancy attracts. Overall, our results show that words in job titles play a crucial role in the first stages of the search and matching process: employers use job titles to advertise their jobs, and workers use job titles to direct their search.[31] The role of higher wages in attracting workers cannot be properly understood without accounting for job titles.

Our results show that job titles are a powerful tool to describe job characteristics, and perform much better than SOC codes across a variety of dimensions. Our findings thus open fruitful avenues for future research to better understand a variety of labor market issues, and in particular human capital investment and the role of management skills in wage differentials.

[31] While we only have data from CareerBuilder, it seems likely that this conclusion holds more broadly. Other employment websites, including Monster.com, Indeed.com and Linkedin.com, use job titles in essentially the same way as CareerBuilder. The evidence in DeVaro & Gürtler (2018) suggests that help-wanted ads in newspaper generally also featured some form of a job title.

References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*. Longitudinal Employer-Household Dynamics Technical Paper 2002-06, Center for Economic Studies, U.S. Census Bureau.

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2), 251–333.

Andrews, M. J., Bradley, S., & Upward, R. (2001). Estimating the probability of a match using microeconomic data for the youth labour market. *Labour Economics*, 8(3), 335–357.

Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 673–697.

Azar, J., Marinescu, I., & Steinbaum, M. I. (2017). *Labor Market Concentration*. Working Paper 24147, National Bureau of Economic Research. DOI: 10.3386/w24147.

Banfi, S. & Villena-Roldan, B. (2018). Do high-wage jobs attract more applicants? directed search evidence from the online labor market. *Journal of Labor Economics*, forthcoming.

Barron, J. M., Bishop, J., & Dunkelberg, W. C. (1985). Employer search: The interviewing and hiring of new employees. *Review of Economics and Statistics*, 67, 43–52.

Barron, J. M., Black, D. A., & Loewenstein, M. A. (1987). Employer size: The implications for search, training, capital investment, starting wages, and wage growth. *Journal of Labor Economics*, 5, 76–89.

Belot, M., Kircher, P., & Muller, P. (2018). How wage announcements affect job search - a field experiment. mimeo.

Bender, S., Bloom, N., Card, D., Van Reenen, J., & Wolter, S. (2017). Management Practices, Workforce Selection, and Productivity. *Journal of Labor Economics*, 36(S1), S371–S409.

Blau, F. D. (1977). *Equal pay in the office*. Lexington, Mass.: Lexington Books.

Blau, F. D. & Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic Perspectives*, 14(4), 75–100.

Bloom, N., Sadun, R., & Reenen, J. V. (2016). *Management as a Technology?* Working Paper 22327, National Bureau of Economic Research. DOI: 10.3386/w22327.

Braun, C., Engelhardt, B., Griffy, B., & Rupert, P. (2016). Do workers direct their search? mimeo.

Brencic, V. (2012). Wage posting: evidence from job ads. *Canadian Journal of Economics*, 45(4), 1529–1559.

Brenčič, V. & Norris, J. B. (2012). Employers' on-line recruitment and screening practices. *Economic Inquiry*, 50(1), 94–111.

Brown, J. & Matsa, D. A. (2016). Boarding a Sinking Ship? An Investigation of Job Applications to Distressed Firms. *The Journal of Finance*, 71(2), 507–550.

CareerBuilder (2013). Job Postings. `http://http://www.careerbuilder.com/jobposter/products/postjobsinfo.aspx`.

comScore Media Metrix (2011). comScore Media Metrix Ranks Top 50 U.S. Web Properties for January 2011. `http://ir.comscore.com/releasedetail.cfm?ReleaseID=551464`.

Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *Quarterly Journal of Economics*, 128(3), 1169–1218.

Davis, S. J. & Marinescu, I. E. (2018). *Posted wages and labor market conditions*. Technical report, working paper.

Davis, S. J. & Samaniego de la Parra, B. (2017). Application flows. mimeo.

DeVaro, J. & Gürtler, O. (2018). Advertising and labor market matching: A tour through the times. *Journal of Labor Economics*, 36(1), 253–307.

Dickens, W. T. & Katz, L. F. (1986). *Interindustry Wage Differences and Industry Characteristics*. Working Paper 2014, National Bureau of Economic Research.

Eeckhout, J. & Kircher, P. (2010). Sorting and decentralized price competition. *Econometrica*, 78, 539–574.

Engelhardt, B. & Rupert, P. (2017). Competitive versus random search with bargaining: An empirical comparison. *Labour Economics*, 48, 183–197.

Faberman, J. R. & Menzio, G. (2017). Evidence on the relationship between recruiting and the starting wage. *Labour Economics*, forthcoming.

Faberman, R. J. & Kudlyak, M. (2014). *The Intensity of Job Search and Search Duration*. Working Paper 14-12, Federal Reserve Bank of Richmond.

Galenianos, M. & Kircher, P. (2009). Directed search with multiple job applications. *Journal of Economic Theory*, 114, 445–471.

Gee, L. K. (2018). The More You Know: Information Effects on Job Application Rates in a Large Field Experiment. *Management Science*.

Gibbons, R. & Katz, L. F. (1992). Does unmeasured ability explain inter-industry wage differentials? *Review of Economic Studies*, 59(3), 515–35.

Groes, F., Manovskii, I., & Kircher, P. (2015). The U-shapes of occupational mobility. *Review of Economic Studies*, forthcoming.

Groshen, E. L. (1991). The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *The Journal of Human Resources*, 26(3), 457–472.

Hall, R. E. & Krueger, A. B. (2012). Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. *American Economic Journal: Macroeconomics*, 4(4), 56–67.

Hershbein, B. & Kahn, L. (2016). Is technological change exacerbated in recessions? evidence from vacancy postings. mimeo.

Holzer, H., Katz, L., & Krueger, A. (1991). Job queues and wages. *Quarterly Journal of Economics*, 106, 739–768.

Hornstein, A., Krusell, P., & Violante, G. L. (2007). *Frictional Wage Dispersion in Search Models: A Quantitative Assessment*. Working Paper 13674, National Bureau of Economic Research.

Iranzo, S., Schivardi, F., & Tosetti, E. (2008). Skill dispersion and firm productivity: An analysis with employer-employee matched data. *Journal of Labor Economics*, 26(2), 247–285.

Kambourov, G. & Manovskii, I. (2009). Occupational mobility and wage inequality. *Review of Economic Studies*, 76, 731–759.

Krueger, A. B. & Summers, L. H. (1986). *Reflections on the Inter-Industry Wage Structure*. Working Paper 1968, National Bureau of Economic Research.

Krueger, A. B. & Summers, L. H. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica*, 56(2), 259–93.

Kudlyak, M., Lkhagvasuren, D., & Sysuyev, R. (2012). Systematic job search: New evidence from individual job application data. mimeo.

Kuhn, P. & Mansour, H. (2014). Is internet job search still ineffective. *Economic Journal*, 124(581), 1213–1233.

Kuhn, P. & Shen, K. (2012). Gender discrimination in job ads: Evidence from china. mimeo.

Lazear, E. P. (1995). A jobs-based analysis of labor markets. *The American Economic Review*, 85(2), 260–265.

Lentz, R., Maibom, J., & Moen, E. R. (2018). Competitive or random search? mimeo.

Manning, A. (2011). Chapter 11 - imperfect competition in the labor market. In *Handbook of Labor Economics*, volume Volume 4, Part B (pp. 973–1041). Elsevier.

Marinescu, I. (2017). The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. *Journal of Public Economics*, 150, 14–29.

Marinescu, I. & Rathelot, R. (2018). Mismatch Unemployment and the Geography of Job Search. *American Economic Journal: Macroeconomics*, 10(3), 42–70.

Mellow, W. & Sider, H. (1983). Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics*, 1(4), pp. 331–344.

Michelacci, C. & Suarez, J. (2006). Incomplete wage posting. *Journal of Political Economy*, 114(6), 1098–1123.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105, 385–411.

Murphy, K. & Topel, R. (1987). Unemployment, risk, and earnings: testing for equalizing wage differences in the labor market. In K. Lang & J. Leonard (Eds.), *Unemployment and the Structure of Labor Markets*. Basil Blackwell Inc.

Nakamura, A., Freeman, R., Pyman, A., & Shaw, K. (2008). Jobs online. In D. Autor (Ed.), *Labor Market Intermediation*. Chicago Press.

Pallais, A. (2012). Inefficient hiring in entry-level labor markets. mimeo.

Petrongolo, B. & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39, 390–431.

Poletaev, M. & Robinson, C. (2008). Human capital specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000. *Journal of Labor Economics*, 26(3), 387–420.

Rogerson, R., Shimer, R., & Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 43(4), 959–988.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5), 996–1025.

Stevenson, B. (2008). The impact of the internet of labor flows. In D. Autor (Ed.), *Labor Market Intermediation*. Chicago Press.

Wolthoff, R. P. (2018). Applications and interviews: Firms' recruiting decisions in a frictional labor market. *Review of Economic Studies*, forthcoming.

Woodcock, S. D. (2007). *Match Effects*. Discussion Paper dp07-13, Department of Economics, Simon Fraser University.

Woodcock, S. D. (2008). Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. *Labour Economics*, 15(4), 771–793.

Yashiv, E. (2007). Labor search and matching in macroeconomics. *European Economic Review*, 51(8), 1859 – 1895.

Table 1: Summary Statistics

	obs.	mean	s.d.	min	max
Job characteristics					
Yearly wage	11,715	57,323	31,690	13,500	185,000
Required experience	168	3.20	2.80	0.50	10.00
Required education	25,931	14.88	1.88	12.00	24.00
Vacancy duration	61,135	15.67	8.86	0.00	31.00
Firm characteristics					
Number of employees	61,135	18,824	59,280	1	2,100,000
Outcome variables					
Number of views	61,135	6,084.02	6,133.50	0	262,160
Number of clicks	61,135	280.97	312.11	0	7,519
Number of applications	61,135	59.35	121.68	0	4,984
Clicks per 100 views	60,979	5.64	5.58	0	100
Applications per 100 views	61,051	1.17	2.57	0	100
Applicant characteristics					
Years of education	2,282	16.63	1.35	14	24
Years of experience	2,379	13.28	5.13	2.5	26

Source: CareerBuilder.com.

Table 2: Ten most common job titles

First 4 words	Freq.	First 2 words	Freq.	First 1 word	Freq.
Panel A: all job titles					
customer service representative	273	customer service	895	senior	4,176
administrative assistant	242	sales representative	865	sales	2,450
project manager	221	director of	580	director	1,080
sales representative	218	project manager	553	customer	1,034
customer service openings in	188	entry level	476	medical	897
sales representative customer service	188	administrative assistant	395	project	883
staff accountant	184	outside sales	374	business	870
outside sales representative	176	inside sales	307	manager	801
senior accountant	166	business development	279	rn	742
full time retail sales	150	business analyst	267	account	654
Panel B: job titles with a posted wage					
First 4 words	Freq.	First 2 words	Freq.	First 1 word	Freq.
customer service representative	120	customer service	260	senior	810
staff accountant	98	administrative assistant	154	sales	369
administrative assistant	93	outside sales	150	customer	277
senior accountant	92	senior accountant	116	administrative	191
executive assistant	63	staff accountant	110	accounting	172
outside sales representative	62	inside sales	106	outside	166
senior financial analyst	56	director of	96	director	146
controller	54	entry level	86	medical	146
financial analyst	49	executive assistant	85	executive	143
chiropractic technician	48	accounts payable	77	account	132

Source: CareerBuilder.com.

Table 3: The impact of wages on the number of applicants per 100 views

	I	II	III	IV	V
Log(Posted Wage)	-0.770***	-0.642***	-0.710***	1.268***	0.947*
	(0.052)	(0.075)	(0.087)	(0.373)	(0.517)
Job characteristics		Yes	Yes		Yes
SOC f.e.		Yes	Yes		
Firm f.e.			Yes		Yes
Job title f.e.				Yes	Yes
Observations	11,708	11,708	11,708	11,708	11,708
R^2	0.017	0.133	0.363	0.464	0.584
Adj. R^2	0.0165	0.0835	0.235	0.123	0.268
AIC	61,152	59,754	57,200	54,049	52,042

Note: The dependent variable is the number of applications per 100 job views. "f.e." stands for "fixed effects". SOC fixed effects are for detailed codes. Job characteristics include vacancy duration, a dummy for salary expressed per hour, required education and experience, designated market area, and calendar month. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: CareerBuilder.com.

Table 4: Words that explain wage residuals after detailed SOC fixed effects

Sign of word coefficient	Job level: seniority / management	Specialization / Skills	Computer
Negative (lower wages)	representative	accountant	network
	assistant	account	
	specialist	project	
	associate	medical	
	entry	marketing	
	coordinator	quality	
	support	inside	
	clerk	bilingual	
	operator	office	
	part-time	advisor	
	staff	receptionist	
	junior	recruiter	
Positive (higher wages)	manager	-	engineer
	senior	sales	developer
	executive	consultant	systems
	director	administrative	software
	management	business	architect
	supervisor	outside	web
	of	with	net
	ii	nurse	java
	lead	maintenance	it
	to	health	
		hr	
		or	
		controller	
		auditor	

Note: The words included are significant at the 5% level in explaining the residuals after a regression of the posted wage on SOC code fixed effects. Words are included when they appear at least 100 times. Words are ordered by frequency and underlined when they appear at least 500 times.

Source: CareerBuilder.com.

Table 5: Explaining applicants' average experience and education

	I	II	III	IV
VARIABLES	Experience (yrs)	Experience (yrs)	Education (yrs)	Education (yrs)
Log(Posted Wage)	2.174***	1.549**	0.757***	0.256**
	(0.203)	(0.657)	(0.045)	(0.130)
Job title f.e.		Yes		Yes
Job characteristics		Yes		Yes
Observations	1,755	1,300	1,696	1,300
R^2	0.238	0.963	0.282	0.976
Adj. R^2	0.238	0.852	0.281	0.900

Note: The dependent variable is the average number of years of experience / education among the applicants to each job. The sample is all jobs with a posted wage. The regressions are weighted by the number of applicants to each vacancy using Stata's analytic weights. Job characteristics include vacancy duration, a dummy for salary expressed per hour, required education and experience, designated market area, and calendar month. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: CareerBuilder.com.

Table 6: The impact of posted wages on clicks per 100 views

	I	II	III	IV	V
Log(Posted Wage)	-1.045***	-0.597***	-0.711***	2.035***	1.930***
	(0.089)	(0.130)	(0.167)	(0.399)	(0.454)
Job characteristics		Yes	Yes		Yes
SOC f.e.		Yes			
Firm f.e.			Yes		Yes
Job title f.e.				Yes	Yes
Observations	11,694	11,694	11,694	11,694	11,694
R^2	0.011	0.168	0.389	0.564	0.643
Adj. R^2	0.011	0.121	0.267	0.287	0.371
AIC	72,956	71,012	68,453	63,366	62,012

Note: The dependent variable is the number of clicks divided by the number of job views divided by 100. "f.e." stands for "fixed effects". SOC fixed effects are for detailed SOC codes. Job characteristics include vacancy duration, a dummy for salary expressed per hour, required education and experience, designated market area, and calendar month. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: CareerBuilder.com.

Table 7: The impact of wages on the number of applicants per 100 views, using occupation weights

	I	II
Log(Posted Wage)	-0.571***	0.833**
	(0.113)	(0.335)
Job characteristics	Yes	
SOC f.e.	Yes	
Job title f.e.		Yes
Observations	11,706	11,706
R-squared	0.198	0.551

Note: The dependent variable is the number of applications per 100 job views. All observations are weighted by occupation-specific pseudo-sampling weights, defined as the ratio of the number of jobs in an SOC in our data to the total number of jobs in that SOC in The Conference Board Help Wanted Online Data Series for January-February 2011. "f.e." stands for "fixed effects". SOC fixed effects are for detailed codes. Job characteristics include vacancy duration, a dummy for salary expressed per hour, required education and experience, designated market area, and calendar month. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: CareerBuilder.com.

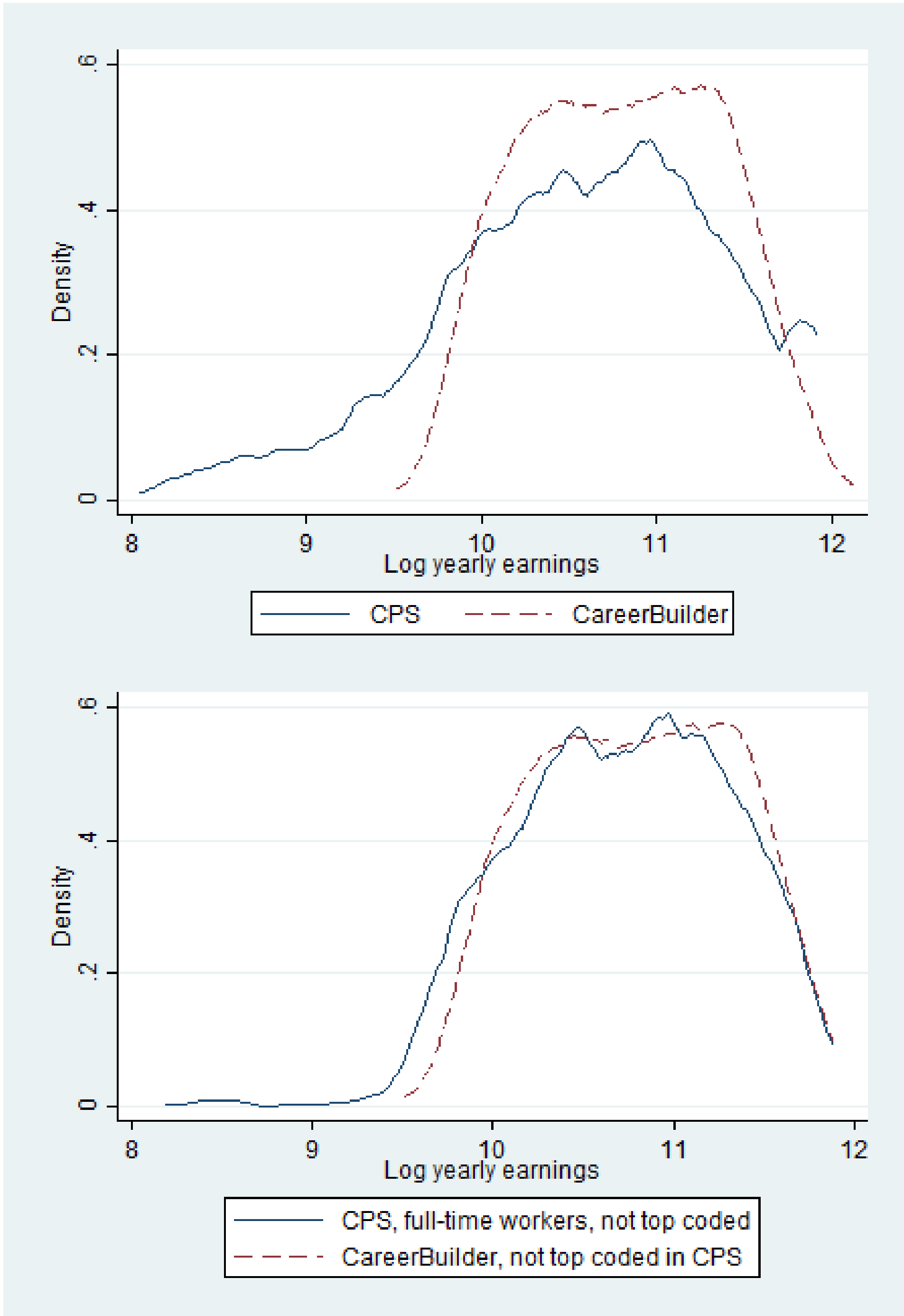
Figure 1: Job titles on CareerBuilder.com

All Jobs:

Jobs with Posted Wage:

Note: Job titles are truncated to the first four words. Only job titles that appear in at least 10 job postings are represented. Word cloud created using www.tagul.com. Tagul uses word frequency to determine the size of the words.
Source: CareerBuilder.com

Figure 2: The distribution of earnings in the CPS vs. posted wages on CareerBuilder.com



Note: A small number of outliers (log yearly earnings < 8) has been omitted for the CPS data.
Source: Current Population Survey and CareerBuilder.com