

# Are Workers Better Matched in Large Labor Markets?

Nikolaj A. Harmon\*  
University of Copenhagen

February, 2013

## Abstract

This paper examines the relationship between labor market size and job search outcomes. Much research and many policy initiatives assume that larger labor markets lead to better job search outcomes because they give workers and firms more choice in potential jobs or employees. The empirical finding that labor market size and job finding rates are uncorrelated, however, has led researchers to question this assumption. I show, theoretically and empirically, that large labor markets may cause workers to find jobs that are better matches given their individual skills and characteristics, even if they do not cause workers to find jobs faster. I construct a unique new data set from Denmark that combines administrative data, an online vacancy database and detailed geographical information. I show that workers in large labor markets find jobs for which they are a better match as measured by both previous industry experience and geographical location. They also find jobs which pay higher wages and result in longer employment spells even after controlling for spatial productivity differences among firms. The estimated effects imply that labor market size explains 6.6% of the spatial variation in wage premia, and also suggest a high rate of return on transport infrastructure projects that increase the effective size of labor markets by increasing workers' ability to commute to distant jobs.

---

\*Email: nikolaj.harmon@econ.ku.dk. I am particularly indebted to David S. Lee, Alex Mas, Stephen J. Redding and Richard Rogerson for many helpful discussions and comments. I also thank Raymond Fisman, Tatiana Homonoff and Rasmus Jørgensen as well as seminar participants at Bocconi University, University of Bonn, Copenhagen Business School, University of Copenhagen, Collegio Carlo Alberto, University of Houston, University of Mannheim, Stockholm IIES, Princeton University and Uppsala University for valuable input. Finally, I would like to thank Mette Birk Kjærgaard and Katarina Ritz at the Danish National Survey and Cadastre, Henning Bunzel at the University of Aarhus as well as Anders Gabriel Pedersen and Claus Bech at the National Labour Market Authority in Denmark for their help with data access. All errors are my own.

# 1 Introduction

A simple intuition suggests that larger labor markets should lead to better job search outcomes: with many vacancies and unemployed workers in the market, it should be more likely that workers can find a suitable job, and firms can find appropriate employees. As a result, much existing research and many popular policy initiatives build on the premise that larger labor markets perform better. In the urban economics literature, for example, the advantages of large labor markets are thought to cause agglomeration and help to explain the existence of cities and industrial clusters like Silicon Valley and San Diego's Biotech Beach.<sup>1</sup> The benefits of large labor markets have also been invoked to justify policies such as transport infrastructure improvements, like bridges or high speed train lines, since they increase workers' ability to commute to jobs further away, which in turn increases the effective size of workers' local labor markets.

The empirical evidence on whether labor market size improves job search outcomes, however, stands in stark contrast to the idea's theoretical popularity. The largest body of evidence comes from the empirical literature on labor market matching functions, which studies the relationship between job finding rates and the total number of unemployed workers and vacancies in the labor market.<sup>2</sup> A consistent finding in this literature is that after controlling for the ratio of vacancies to unemployed workers, job finding rates do not vary with the total number of unemployed workers or vacancies.<sup>3</sup> In the job search literature in particular, this finding has largely been interpreted as evidence that labor market size is unimportant for job search outcomes and has led to a general focus on theoretical models with this feature (Petrongolo and Pissarides (2001)).

By focusing on job finding rates, however, the matching function literature only deals with one aspect of the job search process, namely the speed at which unemployed workers find new jobs. Given the literature's emphasis on understanding movements in and out of

---

<sup>1</sup>See for example Duranton and Puga (2004); Rosenthal and Strange (2004) or Moretti (2011).

<sup>2</sup>An important issue arises in how exactly one defines a labor market. The earliest papers in the matching function literature focus on country-wide labor markets (see for example Pissarides (1986); Blanchard and Diamond. (1990); Layard et al. (1991); van Ours (1991); Burda and Wyplosz. (1994); Berman (1997) and Yashiv (2000)). Positive effects of market size seem less plausible for country-wide labor markets and would imply that larger countries have better functioning labor markets. Later studies, however, also focus on sub-national, local labor markets as defined for example by regions, districts, or so-called travel-to-work-areas (see for example Burda (1993); Bennet and Pinto (1994); Coles and Smith (1996); Munich et al. (1998); Burgess and Profit (2001); Hynninen (2005); Fahr (2006) and Manning and Petrongolo (2011)). It is the effect of market size at this level of aggregation that is the focus of the present paper. Finally, a few papers have also treated different industries or occupations as distinct labor markets (Warren (1996); Anderson and Burgess (2000) and Sahin et al. (2012)). Petrongolo and Pissarides (2001) provide a more detailed review of the literature.

<sup>3</sup>While a few papers do find evidence that market size has a positive effect on job finding rates, a similar number of papers find evidence of a *negative* effect. See the literature review by Petrongolo and Pissarides (2001) as well as the discussion in Petrongolo and Pissarides (2006).

unemployment, this is a natural focus. Looking beyond the determinants of unemployment, however, there could be other important aspects of job search that are affected by labor market size. In particular, this paper examines the possibility that larger labor markets cause workers not to find jobs faster but instead to find jobs that are better matches given their individual skills and other characteristics.

The paper first illustrates theoretically how labor market size could have important effects on worker-job match quality even if job finding rates are unresponsive to market size. Building on a standard two-sided search model, I introduce a positive effect of labor market size on match qualities as well as a type of endogenous separations that allow match qualities to affect employment spell durations. Based on numerical simulations, I then show that the positive effects of market size on match quality may lead to important differences in both wages and employment spell durations across differently sized labor markets. Mirroring a point made previously by Petrongolo and Pissarides (2006), I further show that this can happen even if job finding rates are unresponsive to labor market size. The reason is that workers and firms in larger markets respond by becoming more picky about who they match with which can off-set any positive effects of market size on job finding rates.

I then turn to an empirical examination of the relationship between labor market size and match qualities. In order to conduct a rich analysis of worker-job match quality, I construct a unique new data set containing information on all new hires in Denmark between 2004 and 2006, combining information from a number of sources. Administrative data on workers and firms facilitates the construction of zip code-level data on all individual hires out of unemployment along with detailed characteristics of the matched worker and firm. These data also allow me to measure municipal unemployment stocks. Next, using disaggregated vacancy data from Denmark's largest online vacancy database, *jobnet.dk*, I construct measures of municipal vacancy stocks. Finally, using detailed geographical information, including estimated travel times collected from Google Maps, I merge together the different data sources. I can thus measure the effective size of the local labor market for each new hire by the total number of unemployed workers and vacancies nearby. To my knowledge, this paper is the first to combine zip code level information on the characteristics of new hires with local unemployment and vacancy data.

I begin my empirical analysis by replicating existing results regarding job finding rates and labor market size; workers in larger local labor markets do not find jobs faster. I then look for evidence that the size of the labor market influences the types of matches that occur by regressing the characteristics of new hires on local labor market size. A rich set of worker, firm

and zip code controls are used to alleviate concerns with unobserved differences across labor markets.

I first examine two measures of match quality based on worker and firm characteristics: The first is whether a worker is reemployed in the industry of his previous job. The second is how far a worker must travel to get to his new job. Assuming an unemployed worker has obtained some amount of industry-specific human capital during his prior employment, a worker should be more productive if rehired in the same industry. Similarly, because commuting (or moving) is costly, workers are better off finding a job close to where they live. The empirical results confirm that workers in large markets find jobs that are better matches along these two dimensions: doubling the size of the labor market is estimated to increase the probability that a new hire does not involve a change of industry by 0.5 percentage points (on a base of 18%), while decreasing travel time to the new job by 5.8%.

Motivated by my theoretical analysis, I next examine whether higher match quality in larger labor markets also translates into higher wages and longer employment spells. Regression results show that a doubling of the labor market is associated with 3.2% higher wages among new hires and a 1.6 percentage point increase in the probability that the associated employment spell lasts at least two years (on a base of 44%). Taken as estimates of the causal effect of labor market size, however, these results are likely to be biased. This is because workers in larger labor markets are much more likely to be employed by firms in urban areas, which are widely believed to be more productive for reasons unrelated to labor market size.<sup>4</sup>

To disentangle labor market size effects from the effects of spatial productivity differences among firms, I utilize the detailed geographical information in my data. In particular, I utilize the fact that I can include municipality of work fixed effects in my regressions and still identify the effect of labor market size from the variation in workers' zip code of residence. The inclusion of municipality fixed effects markedly lowers the estimated impact of labor market size on wages and employment spell durations, suggesting that spatial productivity differences are important. The positive effects of labor market size do not disappear however: once spatial productivity differences have been accounted for, doubling the size of the labor market is estimated to lead to 0.6% higher wages among new hires and a 1.0 percentage point increase in the probability that employment spells last at least two years.

Finally, I assess the economic significance of the estimated effects in a couple of ways. First, I quantify the role of labor market size in shaping spatial differences in wage premia. Second, I perform a simple calculation of how the estimated effect of labor market size on

---

<sup>4</sup>See for example Duranton and Puga (2004) and Rosenthal and Strange (2004).

wages affects the profitability of a specific transport infrastructure project: the Bay Bridge in California, which joins together the labor markets of San Francisco and Oakland. Both calculations suggest that the effects labor market size are important in practice: the estimated effect of labor market size on wages can explain about 6.6% of the spatial variation in wage premia, and the simple calculation on the value of the Bay Bridge yields an internal rate of return of 15.4% if the bridge were to be constructed today.

In sum, this paper provides evidence that the size of the labor market has important effects on job search outcomes and in particular on the quality of worker-job matches. This in turn implies that policies which increase effective size of labor markets may have large beneficial effects. The paper's results also suggest that it may be fruitful to devote more attention to job search models with positive effects of labor market size, at least in cases where it is of interest to understand the types of matches that occur and not just the speed at which workers find employment.

While data constraints have limited previous work on the topic, this paper is not the first to look beyond job finding rates in the study of labor market size effects. Petrongolo and Pissarides (2006) were the first to point out how studies looking only at job finding rates may miss important differences in the types of jobs workers find. Studying a single agent search model with an exogenous wage distribution, they show that a positive relationship between labor market size and the distribution of wage offers does not necessarily translate into higher job finding rates if workers in large markets have higher reservation wages. Comparing London to the rest of the UK, Petrongolo and Pissarides (2006) then provide empirical evidence that wage offers are indeed higher in larger labor markets and interpret this as evidence that matches are more productive in larger markets. They are not, however, able to distinguish whether this reflects differences in the types of matches that occur or simply reflects that urban firms have a higher productivity level. Bleakley and Lin (2012) focus more directly on the characteristics of the workers and firms who match. Similar to the present paper's result on industry switching, they study industry and occupation switching across US metropolitan areas and document that unemployed workers in densely populated areas are more likely to be reemployed in the same occupation and industry as their previous job.<sup>5</sup> Since they do not have vacancy data, however, they are unable to relate this to the demand side of local labor markets.<sup>6</sup>

---

<sup>5</sup>Considering also job-to-job transitions, Bleakley and Lin (2012) as well as Wheeler (2008), also show that job changes in densely populated areas are more likely to involve a change of industry and/or occupation for young workers and less likely for older workers. This is consistent with more densely populated areas leading to better job search outcomes also for on-the-job searchers if one assumes that young workers are shopping around and experimenting with different occupations and industries early in their careers, while older worker with more specific human capital prefer staying within the same industry or occupation.

<sup>6</sup>A broader set of papers provide indirect evidence on how labor market size affects job search. An influential

Compared to this existing literature, the present paper makes three major contributions: First, this paper is the first to combine data on unemployment and vacancy stocks with direct measures of match quality based on worker and firm observables. Second, by using geographically disaggregated worker-job data and municipality of work fixed effects, the present paper is also the first paper to disentangle the wage effects of labor market size from the effects of spatial productivity differences among firms. Finally, the paper provides evidence on how labor market size affects commuting times and employment spell durations, both of which have not been studied at all in previous work.

The rest of the paper is organized as follows: Section 2 illustrates theoretically how a positive relationship between labor market size and worker-job match qualities can exist even if job finding rates are unresponsive to market size. Section 3 discusses the construction of the data set on new hires and local labor market in Denmark. Section 4 then presents and discusses the empirical results regarding the effects of labor market size and Section 5 concludes.

## 2 Theoretical analysis

The present paper examines the possibility that the size of the labor market has important implications for the quality of the worker-job matches despite being unimportant for the speed with which workers find jobs. To illustrate theoretically how this can occur, the present section presents a two-sided search model where increases in labor market size cause workers to find jobs that are better matches and in turn yield higher wages and longer employment spell durations. At the same time, however, increases in labor market size does not necessarily lead to faster job finding rates because workers and firms in larger markets respond by becoming more picky in who they match with.

The model I focus on is similar to the standard two-sided search model with heterogeneous match qualities and Nash-bargaining (see for example Mortensen and Pissarides (1999)), however, I modify and extend the model in three ways: First, to capture the idea that workers in good matches may stay employed longer (Jovanovic (1979)) and that market size might therefore affect employment spell durations by affecting the quality of matches (Moretti (2011)),

---

study by Costa and Kahn (2000) documents that highly educated couples have become much more likely to locate in big cities in the US. They interpret this as evidence that improved job search outcomes in larger labor markets helps solve the co-location problem for dual career households. Papageorgiou (2010) analyses a model in which workers in big cities are better able to sort into occupations that fits their skill set. He then confirms the model's empirical predictions in US data. Finally, Andersson et al. (2007) and Mion and Naticchioni (2009) use administrative data to relate the degree of *positive assortative matching* (whether high skill workers are more likely to work for high productivity firms) to local employment density. Andersson et al. (2007) finds that denser areas have more positive assortative matching in the US, however, Mion and Naticchioni (2009) finds that the reverse is true for Italy.

I introduce a simple type of endogenous separations. Second, I assume a search technology that introduces positive effects of market size in a manner consistent with existing empirical evidence. Third, since the aim of the paper is to examine changes in the size of the labor market for a given level of market tightness (a given ratio of unemployed workers to vacant jobs), I treat both the number of jobs and the number of workers in the labor market as exogenous parameters. Doing so makes it straightforward to exogenously vary the market size while keeping the market tightness fixed.<sup>7</sup>

I emphasize at the outset that the point of the theoretical analysis is only to show one example of how positive effects of market size on match quality can be reconciled with the existing evidence regarding job finding rates. In particular, I do not wish to argue that the highly stylized model and search technology considered here is the only one that can generate the patterns found in the data.

## 2.1 Model basics

The model setup is as follows: The economy consists of a fixed measure of  $l$  workers and a fixed measure of  $n$  firms. For ease of exposition, each firm is assumed to have only one job so the terms firm and job can be used interchangeably. Throughout, I shall impose  $n = l$  meaning that the ratio of workers to jobs will remain fixed at one.

Time in the model is continuous and workers and firms are expected income maximizers who all discount at rate  $\rho$ . At each point in time, a worker in the model is either unemployed or employed, while each job is either vacant or filled by an employed worker. I let  $u$  denote the measure of workers who are unemployed and let  $v$  denote the measure of vacant jobs. Unemployed workers receive a flow utility with a monetary value of  $b$ , while vacant jobs incur a flow cost of  $a$ .

A job that is filled by a worker produces a flow output,  $\theta$ , net of any utility or commuting costs. Since not all workers are equally well suited for all jobs, however,  $\theta$  will be assumed to vary across worker-job pairs. In particular, the variation in  $\theta$  can be thought of as reflecting differences in workers' skills and jobs' skill requirements or differences in the commuting distance from a workers's home to a given job. In line with this interpretation, I shall refer to  $\theta$  also as *match quality*.  $\theta$  is assumed to be distributed i.i.d. across worker-firm pairs according

---

<sup>7</sup>If either the number of jobs or workers is instead assumed to vary endogenously with parameters, it is difficult to analyze the effect of market size in isolation because the size of the unemployment and vacancy stocks then only change as other parameters change and does not necessarily change in the same proportion. For the empirical analysis later, however, it is of course important to note that the variation in labor market size in the data may be driven by other differences across markets, such as differences in firm productivity. I return to this in the empirical analysis in section .

to the distribution function  $G$ . Wages for employed workers are set by Nash-bargaining over the flow output from the filled job.

Matched firms and workers are assumed to separate exogenously at rate  $\delta_0$  as usual. However, in addition to these exogenous separation shocks, filled jobs are also assumed to be hit by negative productivity shocks at rate  $\delta_1$ . When a filled job is hit by such a negative shock, the involved worker-firm pair must choose between separating or jointly paying a one-time training/investment cost of  $c$  to keep the match productive. The size of the cost  $c$  is stochastic and drawn from some distribution  $H$ . Worker-firm pairs are assumed to Nash-bargain over whether to pay the cost and how to split it. Note that the introduction of these types of endogenous separations will imply that employment spell durations vary with match quality, as worker-firm pairs with low match qualities will be less likely to pay the training/investment cost when they are hit by a negative productivity shock.

## 2.2 Search and matching

Search and matching in the model is assumed to take place as follows: Firms with a vacant job post a vacancy notice and then waits a period of time,  $\phi$ , before receiving job applications from a number of randomly selected workers,  $z$ . Here  $z$  is a nonzero random variable whose distribution may depend on the number of unemployed workers and vacancies in the market.<sup>8</sup> The assumption of a deterministic waiting period  $\phi$  can be thought of as vacancies posting with an application deadline and workers only applying to a vacancy right at the deadline. The variation in how many workers apply to a given job should be thought of as reflecting randomness in which workers notice a given vacancy and choose to apply for it.

For simplicity, I do not model the details of the worker's application process but simply impose the reduced form assumption that  $z$  is distributed as the strictly positive truncation of a Poisson random variable with mean  $m$ .<sup>9</sup> To capture the fact that the number of applications received may depend on the total number of workers and vacancies in the market, the mean  $m$  is further assumed to depend on the total number of unemployed workers as follows:

$$m = Bu^\eta \left(\frac{v}{u}\right)^\alpha \tag{1}$$

---

<sup>8</sup>The assumption that a vacancy always receives at least one application is clearly unrealistic and is made purely for expositional convenience. All the qualitative conclusions of the model go through if firms have a positive probability of not receiving any applications. In particular, the conclusion that worker's in larger markets do not meet firms noticeably faster than in small markets, goes through as long as the probability of not receiving any applications is sufficiently unresponsive to labor market size.

<sup>9</sup>A more in-depth examination of the individual level job search behavior that gives rise to positive effects of market size would certainly be interesting. Given the empirical focus of the present paper and the fact that the data used later is uninformative about how workers and firms are searching, however, I leave this for future work.



Here  $B$  is simply a scaling parameter, while the parameter  $\alpha$  is seen to govern how the expected number of applications responds to increases in the number of unemployed workers,  $u$ , when the ratio of workers to vacancies,  $\frac{v}{u}$  (the market tightness), is kept fixed. The parameter  $\alpha$  on the other hand governs how the expected number of applications respond to increases in market tightness for a given size of the unemployment stock. Given the assumption that the economy always has the same number of worker's and jobs, the market tightness will identically equal one,<sup>10</sup> and so without loss of generality I simply assume  $\alpha = 0$ . Conversely, I will impose  $\eta > 0$  throughout, implying that increases in the size of the market cause vacant jobs to receive more applications on average. By giving firms more choice in who to hire, it is this increase in the size of applicant pools that will generate a positive effect of market size on job search outcomes.

After receiving applications from  $z$  workers, a firm with a vacant job learns its match quality,  $\theta$ , with each of the  $z$  workers and then meets with the worker who is the best match. Upon meeting, the worker and firm then jointly decide whether to match (whether the worker is hired) based on their match quality. If the worker and firm choose not to match, the worker remains unemployed and the firm posts another vacancy notice and must again wait for a period of time  $\phi$  before receiving a new batch of applications. As usual, the decision of whether to match or not will in equilibrium be governed by a reservation match quality,  $\bar{\theta}$ , such that workers and firms with a match quality  $\theta \geq \bar{\theta}$  will choose to match.

### 2.3 Meeting rates and match qualities

To understand the effects of market size in the model, it is instructive to consider the two main ingredients in the worker's job search: the rate at which a worker meets firms with vacant jobs, which I will denote by  $\lambda_u$ , and the distribution of the match quality,  $\theta$ , conditional on meeting a firm, which I will denote by  $F$ .

To determine the meeting rate, note that if  $v$  vacancies are posted in steady state, the number of vacant jobs that receive applications at a given point in time will simply be  $\frac{v}{\phi}$ . Since each of these firm meets with one worker from their applicant pool,<sup>11</sup> the number of meetings that occur at a point in time is thus also  $\frac{v}{\phi}$ . Dividing the total number of meetings

---

<sup>10</sup>Since the number of employed workers and filled jobs must be equal by assumption, assuming that the total number of workers and jobs is equal will automatically imply that the number of unemployed workers and vacant jobs is also equal.

<sup>11</sup>Note that here I am implicitly assuming that the same worker is never selected to meet with two different firms. This will be true if it is assumed that no two firms ever receive applications from the same worker. Alternatively it might be assumed that workers "hand over" any extraneous meeting opportunities to another worker with the exact same characteristics.

by the total number of unemployed workers then gives the rate at which an unemployed worker meets with a firm:

$$\lambda_u(u, v) = \frac{v}{u} \cdot \frac{1}{\phi} \quad (2)$$

As is clear, the rate at which workers meet firms is thus increasing in the market tightness,  $\frac{v}{u}$ , but for a given level of tightness does not depend on the the total number of workers or firms.

The other key aspect of job search from the worker's perspective is the distribution of match qualities when they meet a firm. If workers met firms completely at random, the distribution of match quality would simply be the distribution of match qualities across worker-firm pairs,  $G$ . Since the search and matching technology described above assumes that firms screen worker applications based on match quality however, this is no longer the case.

To derive the distribution of match quality among worker-firm pairs that meet, consider a firm that has received  $z$  applications. Since firms meet with the best match from their applicant pool, the match quality between this firm and the worker it meets with will be the max over  $z$  draws from the underlying distribution of match quality,  $G$ , and will therefore be distributed according to the cumulative distribution function  $(G(\theta))^z$ . Trivially, a higher  $z$  here implies a better distribution in the sense of first order stochastic dominance; when firms have more applicants to choose from they can expect to find a better match to meet with. Taking expectation over the distribution of  $z$  yields the following distribution of match qualities among all worker-firm pairs who meet:

$$F(\theta; u, v) = \sum_{i=1}^{\infty} P(z = i; u, v) G(\theta)^i \quad (3)$$

Note that because the distribution of  $z$  depends on the number of unemployed workers and vacancies,  $u$  and  $v$ , so does the distribution of match qualities,  $F(\cdot; u, v)$ . In particular, plugging in for the assumed distribution of  $z$  yields:

$$F(\theta; u, v) = \sum_{i=1}^{\infty} \frac{(Bu^\eta)^i}{(i!) (\exp(Bu^\eta) - 1)} (G(\theta))^i \quad (4)$$

For  $\eta > 0$  it is easily verified that increasing the size of the labor market size (which implies increasing  $u$ ), will lead to a better  $F$  distribution in the sense of first order stochastic dominance. In light of the discussion above, the intuition for this result is simple: because firms with vacant jobs end up with a larger pool of applicants in larger labor markets, the workers they choose to meet with will generally be better matches.

In sum, the search technology described above do not imply that workers in larger markets meet firms faster, however, when workers do meet with a firm, it is more likely to be a good match when the market is large. This is somewhat different from the way market size effects have been modeled in most of the existing literature which has generally entailed having higher meeting rates for workers in larger markets. By deviating from this approach, however, the search process in the present model is more consistent with the empirical evidence in Petrongolo and Pissarides (2006) which suggests that workers in larger markets face a better wage distribution but do not have a higher arrival rate of potential jobs.<sup>12</sup>

## 2.4 Steady state equilibrium

To write down the equations characterizing the model's steady state equilibrium, I need to introduce some additional notation:  $U$  will denote the value of unemployment and  $V_0$  will denote the value of a vacant job at the time of posting a vacancy notice.  $S(\theta)$  will be the value of the match surplus from a type  $\theta$  match.  $w(\theta)$  will be the wage paid to a worker in a type  $\theta$  match. Finally  $K$  will be the steady state distribution of match quality among filled jobs. With this notation in place, the model's steady state equilibrium can then be characterized by equation (2) and (4) above as well as the following additional equations (see Appendix A for details and interpretations):

$$\rho U = b + \lambda_u(u, v) \int_{\bar{\theta}}^{\infty} \beta S(\theta) dF(\theta; u, v) \quad (5)$$

$$V_0 = \int_0^{\phi} -a e^{-\rho t} dt + e^{-\rho \phi} \left( \int_{\bar{\theta}}^{\infty} (1 - \beta) S(\theta) dF(\theta; u, v) + V_0 \right) \quad (6)$$

$$\rho S(\theta) = \theta - \rho U - \rho V_0 + \delta_0 (-S(\theta)) + \delta_1 \int_0^{\infty} \max\{-c, -S(\theta)\} dH(c) \quad (7)$$

$$\bar{\theta} = \rho U + \rho V_0 \quad (8)$$

$$w(\theta) = \beta \theta + (1 - \beta) \rho U \quad (9)$$

$$1 = \frac{(l - u) \int_{\bar{\theta}}^{\theta} (\delta_0 + \delta_1 (1 - H(S(\theta')))) dK(\theta')}{u \lambda_u(u, v) (F(\theta; u, v) - F(\bar{\theta}; u, v))} \quad (10)$$

$$v = l - u - n \quad (11)$$

## 2.5 The effect of market size on equilibrium outcomes

I now turn to an examination of how labor market outcomes vary with market size in the model described above. Due to the introduction of endogenous separations and the resulting

---

<sup>12</sup>Since wages in the present model will be increasing in match quality, facing a better match quality distribution is equivalent to facing higher wage jobs.

nonstandard form of the functional equation for the surplus from a filled job,  $S(\cdot)$ , analytical results are not readily available. Numerical model solutions are available using methods for functional equations, however, and I will therefore illustrate the effect of market size by showing numerical solutions for different levels of market size.

Table 1 shows the parameter values I employ in the numerical simulation. Since the aim of the numerical exercise is only to illustrate how labor market size may simultaneously be important for match qualities but unimportant for job finding rates, I choose a set of parameter values that emphasizes this possibility.<sup>13</sup> In terms of the specific parameter values, note in particular that by choosing the  $\eta$ -parameter to be positive, I am assuming that vacant jobs can expect to receive more applications in larger labor markets. As discussed above, this will imply that match quality will be higher among those worker-firm pairs that meet in larger markets.

In line with previous work (and in line with the assumed search technology), I focus on the equilibrium number of unemployed workers and vacancies,  $u$  and  $v$ , as the relevant measure of labor market size. By solving the model for various levels of the total number of workers and jobs,  $l$  and  $n$ , but imposing  $l = n$  throughout, I am able to induce variation in  $u$  and  $v$ , while keeping the labor market tightness,  $\frac{v}{u}$ , fixed at one. The numerical solution is carried out using a projection method based on Chebyshev collocation. Appendix B goes through the details of the solution method. In terms of equilibrium outcomes, I focus particularly on how labor market size affects the following outcomes: the job finding rate, the average match quality among new hires, the average wage among new hires and the average separation rate among new hires.

The numerical results are shown in Figure 1, which plots the four outcomes from the model against the equilibrium market size as measured by the total number of unemployed workers or vacancies (recall that these are always equal under the parameters considered). To facilitate comparison across the four outcomes, I scale them all to be 100 in the baseline where  $l, n = 1$  and  $u, v = 0.023$ .

Starting with the average match quality among new hires, Figure 1 shows that as market size increases, match qualities improve markedly. Increasing the size of the labor market by 50% is thus seen to increase the average match quality by about 4%. A similarly sized positive

---

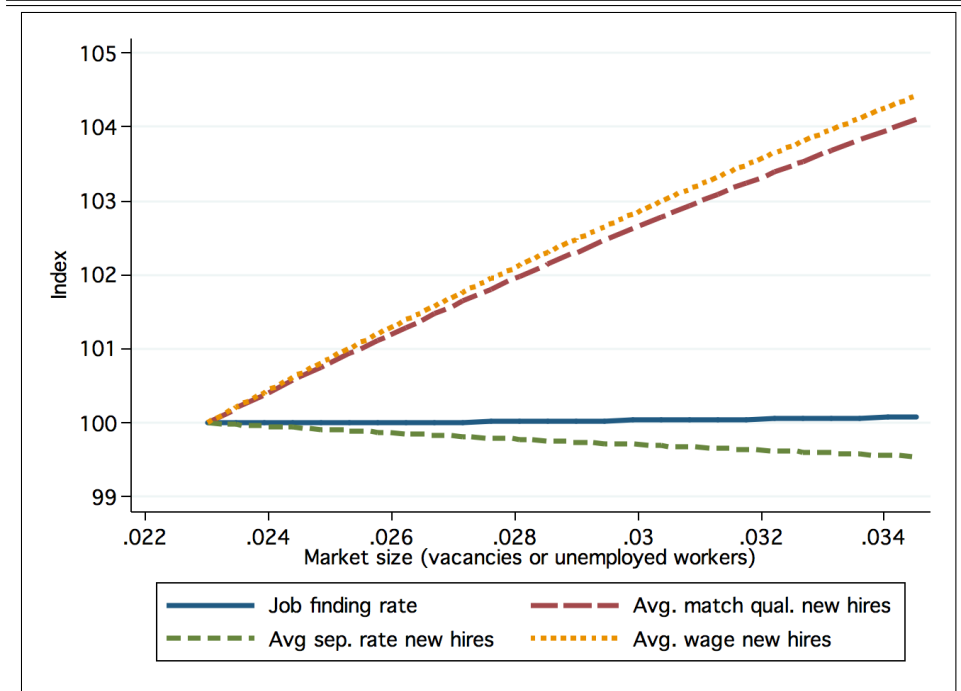
<sup>13</sup>I thus do not calibrate the model to any specific features of the data but postpone empirical testing of the model's predictions to the regression analysis in Section 4. Obviously, an explicit calibration or estimation of the model presented here could provide another empirical test of the effects of labor market size. Such a test, however, would invariably be a test not just of the relationship between market size and search outcomes but of the presented search model as a whole. At the same time, it is well known that stylized search models like the one presented here have difficulties in quantitatively matching many features of labor market data (see for example Shimer (2005) and Hornstein et al. (2011)).

Table 1: Choice of parameters and distributions for numerical analysis

Discount rate:	$\rho = 0.050$
Flow utility, unemployed:	$b = -2.100$
Flow cost, vacancies:	$a = 2.000$
Exogenous separation rate:	$\delta_0 = 0.001$
Arrival rate, negative shocks:	$\delta_1 = 0.145$
Worker bargaining power:	$\beta = 0.500$
Vacancy waiting time:	$\phi = 0.083$
Exp. applications, scaling parameter:	$B = 20.00$
Exp. applications, effect of market size:	$\eta = 0.500$
Match quality distribution:	$G \sim \ln \mathcal{N}(-0.5, 0.6)$
Distribution of negative shocks:	$H \sim \text{Beta}(1, 3)$ on $[0, 30]$

The table lists the parameter values and distributions used in the numerical analysis of the effects of market size. The Beta-distribution used is scaled so that it has positive support on the listed interval. In the numerical work the Lognormal-distribution is truncated from above at 5. The probability mass above 5 in the employed Lognormal-distribution is less than 0.0003.

Figure 1: Numerical results on the effect of market size



The figure plot equilibrium outcomes from the model for different levels of the total number of workers and jobs,  $l$  and  $n$ . The  $x$ -axis measures the equilibrium market size as measured by the total number of vacancies or unemployed workers. The  $y$ -axis measures the different outcomes, scaled so that there are all 100 in the baseline case where  $l, n = 1$  and  $u, v = 0.023$ .

effect of market size is apparent for the average wage among new hires. The effect of labor market size on the average separation rate among new hires is smaller but still noticeable, with a 50% increase in labor market size leading to a drop in the average separation rate of about 0.5%. Looking finally at the job finding rate, however, we see that this barely changes at all as the market size increases. After a 50% increase in the size of the labor market, the job finding rate has increased by less than 0.1%.

Since as discussed above, larger markets lead to better match qualities among those worker-firm pairs that meet, it is quite intuitive that larger labor markets should lead to increases in the equilibrium level of match quality among new hires. Similarly, since wages are increasing in match quality and the separation rate is decreasing, it is in turn also intuitive that larger markets should translate into higher wages and lower separation rates. The fact that job finding rates barely increase at all in larger markets, however, might at first appear somewhat puzzling: if workers in larger markets are more likely to meet firms which are good matches, should that not imply that they find a job significantly faster? To answer this question, start by noting that a worker's job finding rate is simply equal to the rate at which he meets a vacant job times the probability that the the match quality with the job is high enough to result in a match (is above the reservation match quality,  $\bar{\theta}$ ):

$$\lambda_u(u, v) (1 - F(\bar{\theta}; u, v))$$

Now as discussed in Section 2.3 above,  $\lambda_u(u, v)$  does not change with labor market size so the only way in which labor market size affects the job finding rate is by changing the probability of receiving a match quality draw above the reservation output,  $(1 - F(\bar{\theta}; u, v))$ . Since increases in market size cause the distribution of match qualities  $F$  to shift up, it is therefore clear that for a given level of  $\bar{\theta}$ , increases in labor market size will increase the job finding rate. As the labor market size changes, however,  $\bar{\theta}$  will generally not remain fixed but will shift up as workers and firms respond to the upward shift in the  $F$ -distribution and become more picky in who they match with. The net effect of an increase in market size on job finding rates depends on how much of the positive effect of a better match quality distribution is offset by increases in  $\bar{\theta}$ .<sup>14</sup>

Table 2 examines this numerically by showing the effects of increasing the market size by 50% in the model. Column 1 of the table shows the actual effect of such an increase in market

---

<sup>14</sup>It is theoretically possible that reservation output increases so much in response to improvements in the match quality distribution that the net effect on the job finding rate is negative, however, this does not happen under the parameters considered here.

size on the job finding rate, which as before is seen to be less than 0.1%. Column 2 hints at why the job finding rate only changes so little: after the labor market increase, the reservation output has increased 5%. Finally, Column 3 computes the hypothetical increase in the job finding rate that would have occurred if the reservation output had remained fixed across the change in market size. In this case the job finding rate would have increase by a substantial amount, 17.67%. Thus the reason that the job finding rate changes so little in the numerical example in Figure 1 is indeed that as market size goes up, so does the reservation output.

Table 2: The effect of a 50% increase in market size

Actual change in job finding rate	Actual change in $\bar{\theta}$	Hypothetical change in job finding rate with fixed $\bar{\theta}$
0.08%	5.00%	17.67%

The table shows the effects of a 50% increase in market size over the baseline market size where  $l, n = 1$  and  $u, v = 0.023$ . The first two columns shows the percentage change in the equilibrium job finding rate along with the percentage change in the reservation output,  $\bar{\theta}$ . The third column shows the change in in the job finding rate that would have occurred if the reservation output  $\bar{\theta}$  had been assumed to remain fixed across the increase in market size.

Summing up, the two-sided search model presented here yields two main take-aways regarding the effect of labor market size on job search: First, by increasing the quality of the worker-job matches that occur, labor market size can have important effects on both wages and employment spell durations. Second, and more interesting, this can happen even if job finding rates are unresponsive to labor market size. Despite the existing empirical evidence that labor market size is unimportant for job finding rates, labor market size could thus still have important implications for the the types matches that occur. The rest of this paper therefore provides empirical evidence on the effects of labor market size, focusing in particular on worker-job match quality and its effect on wages and employment spell durations.

### 3 Data

To examine the relationship between labor market size and the quality of the worker-job matches that occur, I construct a unique new data set on all hires across local labor markets in Denmark from 2004 to 2006. I construct the final data set by separately constructing data on new hires, unemployment stocks and vacancy stocks and then combining these using detailed

geographical information. Below I go through each of these steps in turn.

### 3.1 New hires

For the data on new hires, I rely on the Danish IDA database, which contains administrative data on all workers, firms<sup>15</sup> and employment relationships in Denmark. From the data on employment relationships as well data on worker’s unemployment spells I construct a data set containing all new hires out of unemployment from 2004 to 2006.<sup>16</sup> I define a new hire out of unemployment<sup>17</sup> to be an employment relationship preceded by at least three weeks of unemployment on the part of the worker and involving a worker-firm pair that has not had any other employment relationships over the last two calendar years. One limitation of the data I am using is that they do not contain information on the within-year timing of either employment relationships or unemployment spells. In constructing the data on new hires, I therefore have to make assumptions about the timing of events in cases where this is ambiguous. Appendix D presents the details of how this is done. In addition, the lack of within-year timing information requires me to aggregate the time dimension to yearly levels.

Based on the worker and firm identifiers in the data, I can merge in detailed information on the worker-firm pair involved in each new hire. In doing so, I am able to tie each hire to a geographical location based on the involved worker’s zip code of residence.<sup>18</sup> In addition, I also observe the municipality of the hiring firm. On the worker side I include a set of basic ”Mincer” controls: years of previous work experience, years of schooling and gender. I also include a variable detailing exactly what type of education the worker has received. This variable groups a worker’s education into 511 different categories (344 of which actually show up in the final data) based on the highest completed education (including apprenticeship degrees and vocational training).<sup>19</sup> Finally, I include the firm industry.<sup>20</sup>

---

<sup>15</sup>To be more precise, the IDA database both contains data on individual establishments as well as the groups of establishments which together make up firms. With the exception of Appendix D, I will only be using establishment-level data. I shall, however, simply refer to the establishment-level variables as firm variables.

<sup>16</sup>I focus on unemployed job search and exclude job-to-job transitions because the empirical implications of market size effects are clearer for unemployed search and hence are easier to test. Consider for example the rate at which employed workers make job-to-job transitions. If larger labor markets positively affects job search, the rate of job-to-job transitions could be higher in large markets because on-the-job searchers have an easier time finding better jobs. On the other hand, employed workers in larger labor markets would also be more likely to already be in a good match, which could makes it less likely that they make a job-to-job transition. Papageorgiou (2010) provides a more formal discussion of this ambiguity in the context of occupational switching.

<sup>17</sup>A worker is measured to be unemployed in IDA if he is receiving some form of UI (*dagpenge* or *kontanthjælp*)

<sup>18</sup>The zip code of residence information is as of January 1st so there is likely to be some measurement error in the geographical location as workers may have moved in between January 1st and the date at which they are hired.

<sup>19</sup>Examples of such education categories are *Masters degree in business economics and management* or *Apprenticeship degree in construction*.

<sup>20</sup>The involved industry classification is unique to Denmark but is based on the Danish version of the EU’s NACE classification. In terms of aggregation-level it is somewhat more detailed than 2-digit industries in NACE but less



Since the data allow me to track workers over time, I next determine for each hire, whether the involved worker is being reemployed in the same industry as his previous job. Similarly, I can determine the length of the employment spell associated with each hire. In particular, I use data up to 2008 and track whether the hired worker becomes unemployed again<sup>21</sup> to construct an indicator for whether each hire results in an employment spell of at least two years.

The employment relationship data contain information both on the number of days the worker was employed within each year and the total wages paid. I use these to construct a measure of the starting monthly wage by dividing total wages by the number of days of employment in the first year. To deal with the potential outliers created by this process, I trim the top and bottom half percent of the monthly wages.

Finally, I drop hires where the worker's experience or education data is missing or where the worker is currently a student. I also drop hires that involve a few very small zip codes and municipalities on smaller islands.<sup>22</sup>

### 3.2 Data on unemployment stocks

For the data on unemployment stocks, I again rely on the IDA database. By knowing the share of the year that each worker spends in unemployment,<sup>23</sup> I construct the average number of unemployed workers in each municipality by summing this share over all workers living in the municipality.<sup>24</sup> As above, I drop students.

### 3.3 Vacancy stocks

To construct geographically disaggregated data on vacancies, I obtain data on job listings from *jobnet.dk*, a government-run online vacancy database launched in 2004. The data contains information on the daily average number of vacancies in each municipality for each month since 2004. From this I drop apprenticeship positions and then take yearly averages to obtain data on the average number of vacancies in each municipality in each year 2004-2006.

---

detailed than 3-digit industries.

<sup>21</sup>As in the construction of the hiring data, I treat unemployment spells of less than three weeks as job-to-job transitions

<sup>22</sup>I drop all hires from zip codes on the small island of Bornholm, which is isolated far east of the rest of Denmark. I also drop all hires from zip codes on very small islands (Læsø, Fanø, Samsø, Anholt, Sejerø, Fur and Fejø) and from one other very small zip code (Vejers Strand). Finally I drop all hires involving zip codes or municipalities on the small island of Ærø, which experienced a change in the municipal structure during the sample period.

<sup>23</sup>A worker is measured to be unemployed in IDA if he is receiving some form of UI (*dagpenge* or *kontanthjælp*).

<sup>24</sup>Of course, since the data also contains workers' zip code of residence, it would also be possible to construct vacancy stocks at the zip code level, however since the vacancy data is only available at the municipal-level, I construct the unemployment stocks at the municipal-level as well.

With the exception of the concurrent work by Sahin et al. (2012) on the HWOL database for the US, I am unaware of other papers using or documenting vacancy data from online databases. The representativeness and coverage of the *jobnet.dk* data therefore warrants some discussion.

A priori, there is good reason to expect that the data are representative and have reasonable coverage: internet usage in general is high in Denmark<sup>25</sup> and *jobnet.dk* is both the largest online vacancy database in Denmark and also exchanges vacancy info with other databases. In addition, all public sector vacancies are required to be listed on *jobnet.dk*.

To get a better sense of the data's coverage and representativeness, I compare it to the official vacancy measure from Statistics Denmark, which is based on a survey of employers and is constructed to be representative of the entire Danish labor market. Figure 2 plots vacancy data from *jobnet.dk* against the official, survey-based measure for the years when this is available (2010-2011). Each point in the figure corresponds to a quarter and a major geographical region,<sup>26</sup> since this is the finest disaggregation at which the survey-based measure is available. The line in the figure is the corresponding regression line. The figure suggests that the vacancy data from *jobnet.dk* follows the official survey-based data closely. All the points are clustered tightly around the regression line, reflecting that the correlation between the two measure is 0.95. There is, however, clear evidence that the vacancy data does not capture all vacancies in the economy, with the survey-based measure being between 2.5 and 3 times higher than the *jobnet.dk* measure.

Figure 3 provides another check on the data by plotting the country-wide average number of vacancies for each month between January 2004 and December 2006. The figure shows that *jobnet.dk* usage has increased systematically across the sample period, with vacancies showing a positive trend that is too big to be explained by real fluctuations in vacancies. Throughout the empirical analysis, I control for time fixed effects. This picks up the time trend in usage as long as it is similar across areas.

### 3.4 Geographical information

Panel A and B of Figure 4 shows maps of the Danish zip codes and municipalities, respectively<sup>27</sup> and plots their centroids. Figure 5 combines the two sets of centroids on one larger map, with

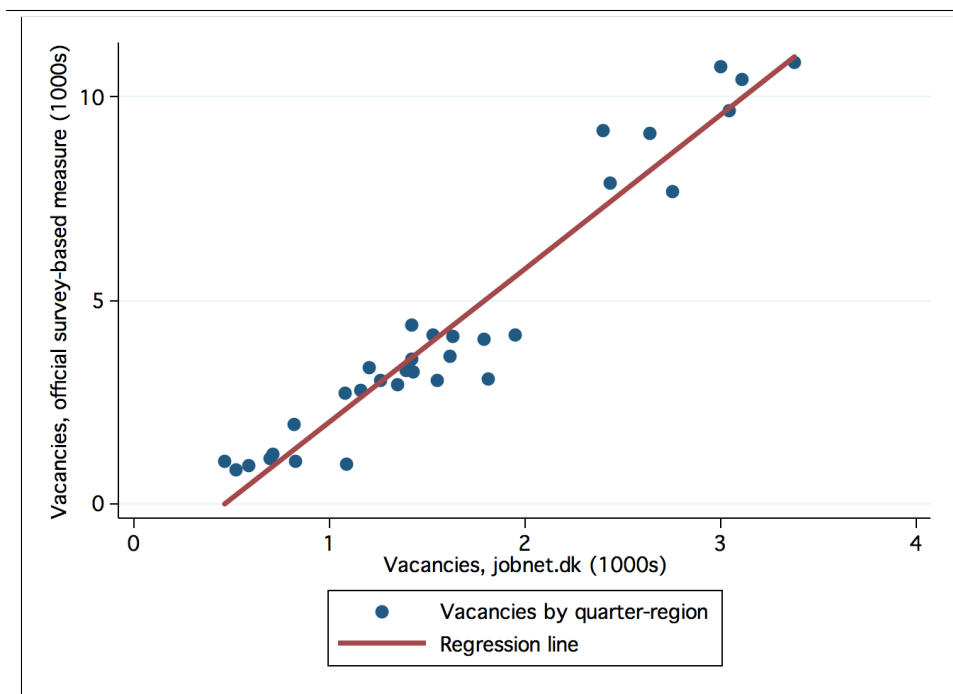
---

<sup>25</sup>According to Statistics Denmark, for example, 71% of households had internet access in Denmark in 2004 and the percentage is likely significantly bigger if restricted to the working age population. Data from a CPS supplement on internet and computer use puts the similar figure for the US at 55%.

<sup>26</sup>The four regions are: Nordjylland, Midtjylland, Syddanmark and Sjælland.

<sup>27</sup>The municipalities and zip codes on the island of Bornholm are not shown in the figure. These are also not included in the empirical analysis cf. footnote 22.

Figure 2: Comparing survey-based and jobnet.dk vacancy measures



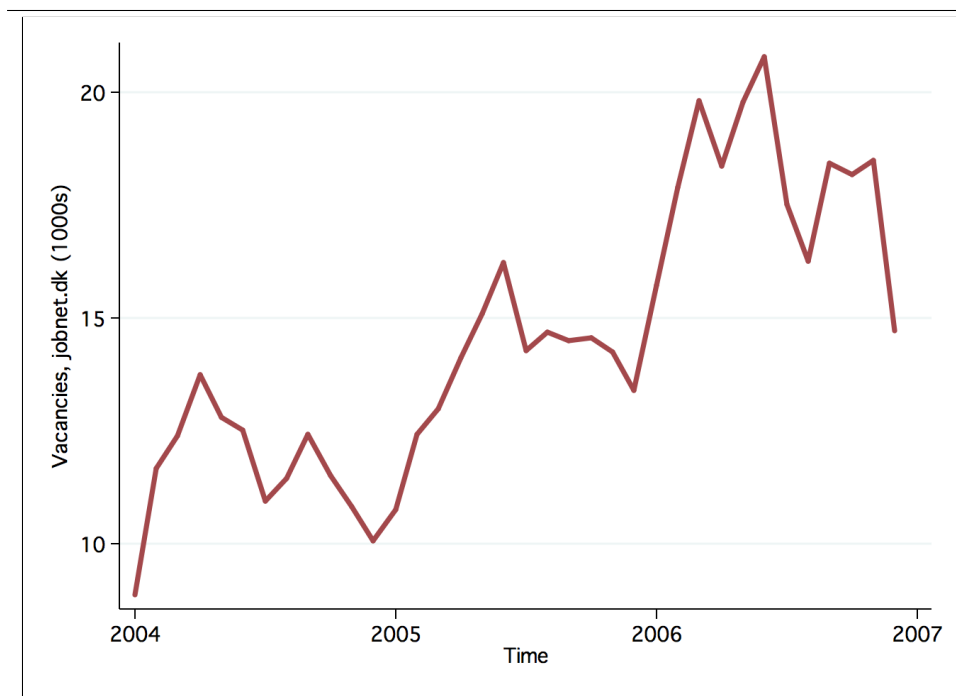
The figure plots the daily average number of vacancies according to Statistics Denmark’s official, survey-based measure against the the daily average number of vacancies in the *jobnet.dk* data. The level of observations are region by year for each of four major regions in Denmark and each quarter 2010-2011. The red line corresponds to an OLS regression and has a slope of 3.76 and a constant term of -1.76.

red dots denoting municipality centroids and blue dots denoting zip code centroids. The data sets described in the preceding sections thus contain information on unemployed workers and vacancies at each of the red dots in Figure 5, while each new hire in my data can be tied to one of the blue dots based on the hired worker’s zip code of residence.

In order to relate the information on new hires to the size of the local labor market, I need to define which workers and vacancies are part of the local labor market where each of the new hires took place. The most common way of defining local labor markets in empirical work has been to simply assume that local labor markets follow some set of statistical or administrative boundaries dictated largely by the available data. Besides ignoring the fact that real world labor markets overlap,<sup>28</sup> this definition would in the present data either define local labor markets as implausibly narrow or would completely obscure the data’s detailed geographical disaggregation. Instead of basing the definition of local labor markets on administrative boundaries, I will therefore define the local labor market for each worker to consist of all the

<sup>28</sup>Manning and Petrongolo (2011) and Lottmann (2012), for example, show that overlapping labor markets is an important feature of the job search process.

Figure 3: Total jobnet.dk vacancies over time



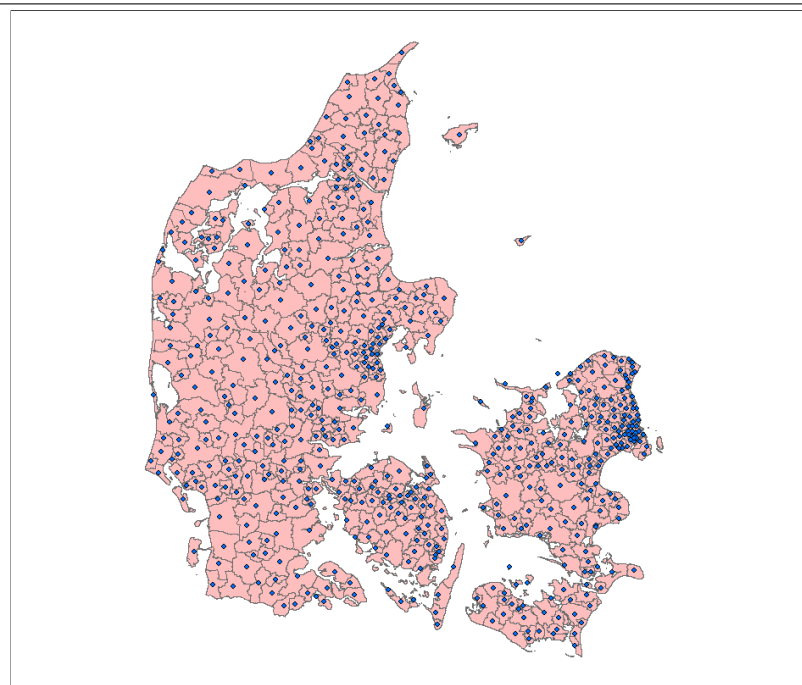
The figure plots the daily average number of vacancies throughout Denmark according to the *jobnet.dk* data for each month 2004-2006.

unemployed workers and vacancies that are sufficiently close to the centroid of his zip code of residence. In order to do this, however, I of course need a measure of closeness.

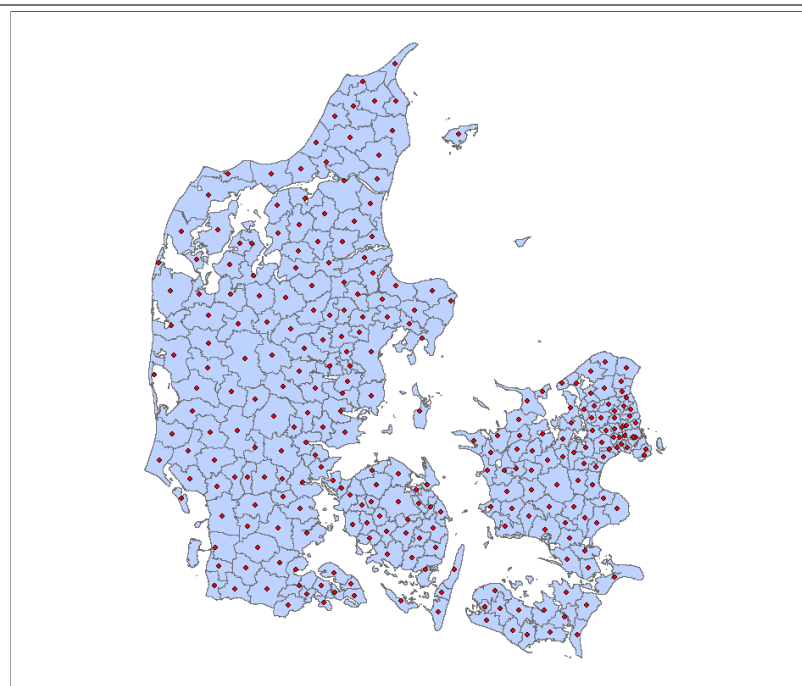
Figure 6 shows two ways of measuring distance between municipality and zip code centroids using zip code Gørlev and Kerteminde municipality as examples. Panel A considers the Euclidean distance, which is seen to be about 35 kilometers. The panel also shows, however, that for a worker in Gørlev to travel only 35 kilometers and get to a vacancy in Kerteminde, he would have to travel most of the way by sea while crossing a body of water known as the Great Belt. In panel B, I instead show the driving directions produced by Google Maps if one asks for directions from Gørlev's centroid to Kerteminde's centroid. Since these take into account the fact that a worker traveling from Gørlev to Kerteminde would have to first travel south to cross the Great Belt Bridge, the resulting travel distance is 78 kilometers, more than twice the Euclidean distance. As this example shows, the Euclidean distance will not be a good approximation of proximity in a country with as rich a geography as Denmark. More importantly, one advantage of studying the effects of labor market size in a Danish context is exactly that the rich Danish geography creates valuable variation in labor market size. This variation is largely lost, however, if distances are measured using the Euclidean distance.

Figure 4: Danish zip codes and Municipalities

*Panel A: The Danish zip codes*

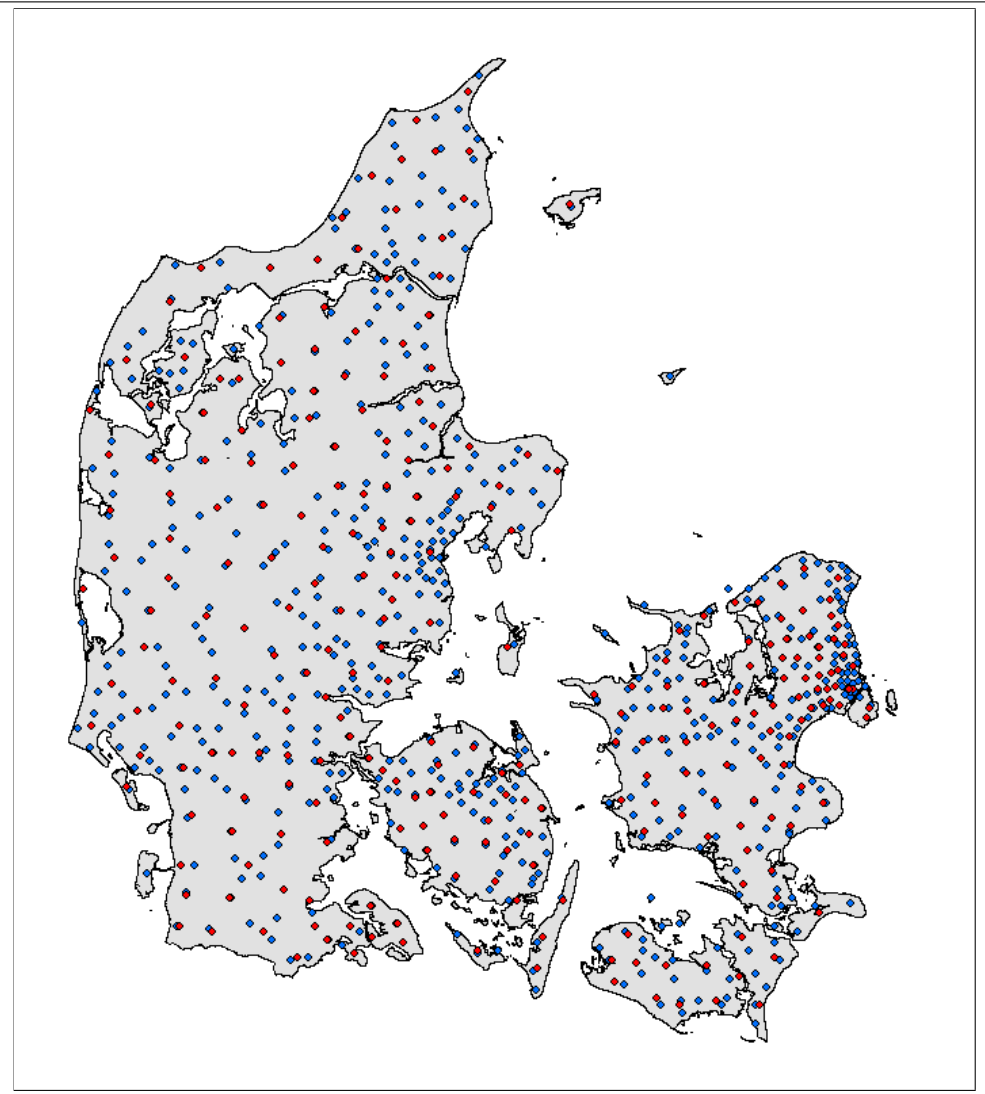


*Panel B: The Danish municipalities*



The two panels show maps of the Danish zip codes and municipalities, excluding the island of Bornholm. The dots show zip code and municipality centroids.

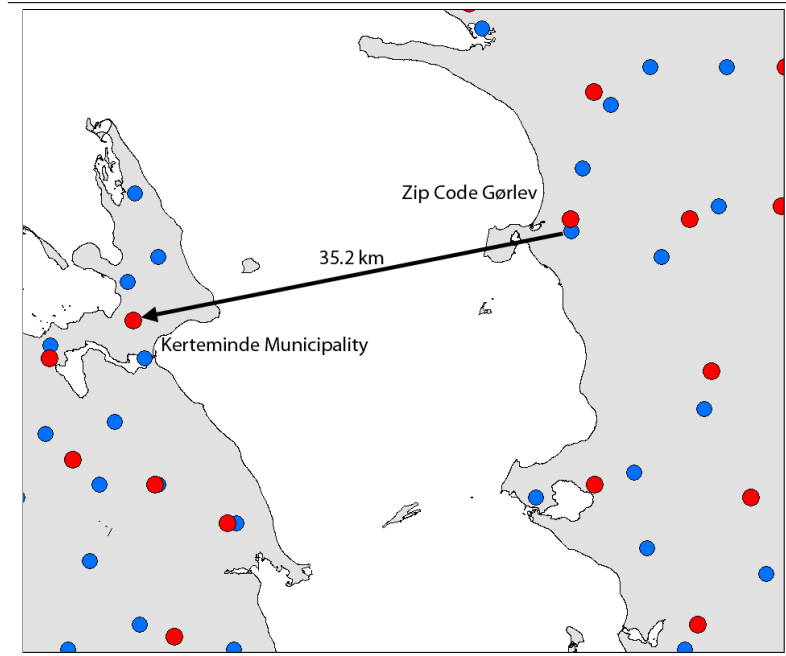
Figure 5: Centroids of the Danish zip codes and municipalities



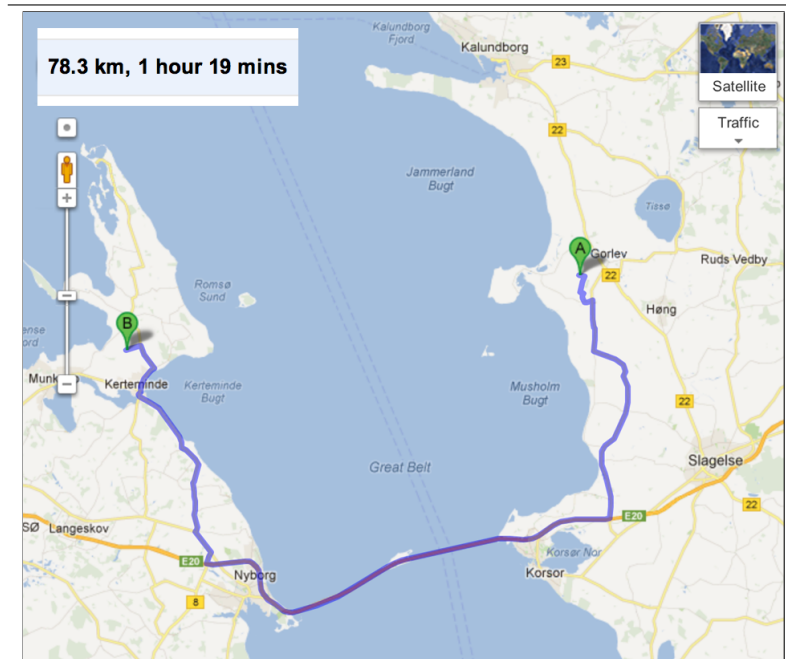
The figure shows a map of Denmark, excluding the island of Bornholm. The blue dots correspond to zip code centroids and the red dots correspond to municipality centroids.

Figure 6: Measuring the distances between centroids

*Panel A: Euclidean distance*



*Panel B: Google Maps driving directions*



Panel A and B shows maps of the Great Belt region in Denmark. The arrowed line in Panel A corresponds to the Euclidean distance between the centroids of zip code Gørlev and Kerteminde municipality. Panel B shows Google Maps driving directions from the centroid of zip code Gørlev to the centroid of Kerteminde municipality.

To construct a measure of distance that accounts for the rich Danish geography, I collect driving directions from Google Maps for all of the centroid pairs in Figure 5 and use the estimated travel times to measure distances between centroid pairs. In addition to dealing with geographical features, the use of estimated travel time also takes into differences in road speeds.<sup>29</sup>

Finally, I collect data on two zip code characteristics from the online Statistikbanken database at Statistics Denmark: total population and total land area. These variables allow me to control for local population density in the empirical analysis.

### 3.5 Measuring local labor markets

From the zip code and municipality data and the geographical distance data described in the previous section I now construct data on the size of the local labor market for each zip code and year. For some given  $t$ , I define the local labor market for workers in a zip code as consisting of all the unemployed workers and vacancies that are within  $t$  minutes of the zip code centroid. I compute the number of unemployed workers and vacancies in each local labor market by assuming that unemployed workers and vacancies in municipalities are distributed uniformly on a disc around their municipal centroid.<sup>30</sup>

In the empirical analysis later, I choose  $t$  equal to 60, implying that local labor markets consist of all unemployed workers and vacancies within one hour. Because the data on new hires contain information on the hired workers zip code of residence and his municipality of work, I can use the travel time data discussed above to measure each hired worker's travel time to his new job. To examine whether one hour is a reasonable cut-off for measuring local labor markets, Figure 7 thus shows the distribution of travel times among the sample of new hires. As is clear from the figure, the vast majority of hires involve workers and firms located less than one hour from each other. More precisely, 92% of all new hires involve a travel time of less than one hour. This suggests that  $t = 60$  does a good job of capturing the effective size of workers' local labor markets.<sup>31</sup>

---

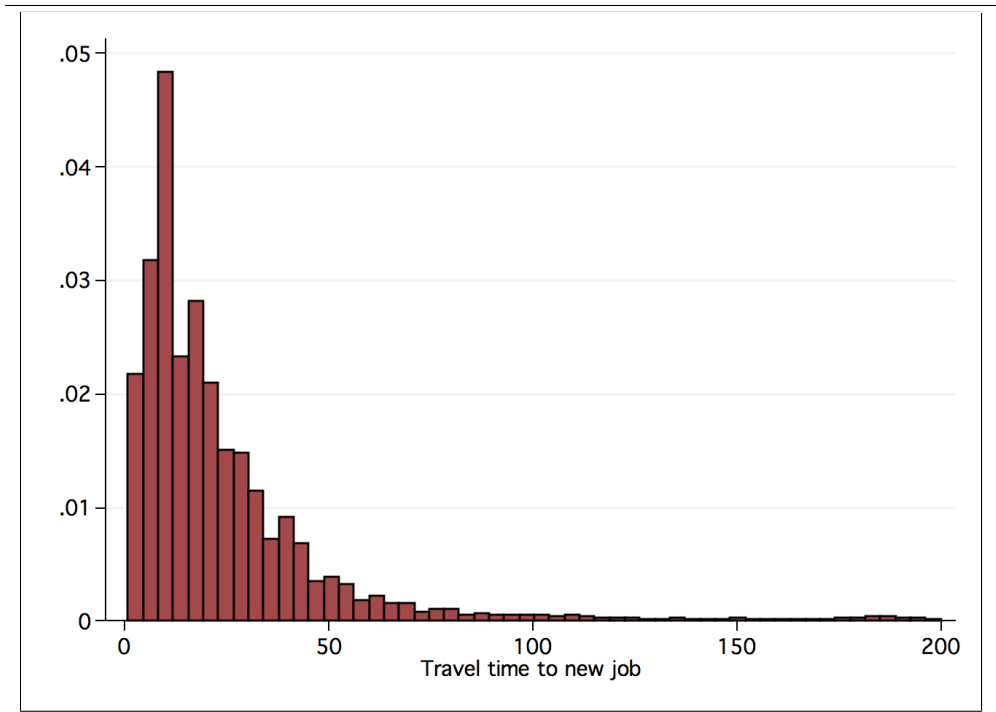
<sup>29</sup>The fact that the Google travel times used does not factor in congestion and the fact that not all workers commute by car, introduces some measurement error. This can be expected to weaken results slightly.

<sup>30</sup>The measured size of the labor market becomes discontinuous in  $t$  and highly sensitive to the choice of  $t$  if one simply attributes all or none of the workers in a municipality to a local labor market depending on whether the municipality centroid is within  $t$  minutes of the zip code centroid. The assumption that vacancies and unemployed workers are distributed on a disc around the municipality centroid ensures that the computed labor market size is a continuous function of  $t$ . Appendix E presents the details.

<sup>31</sup>Experimenting with different cut-offs in the empirical analysis yields results that are very similar to the ones presented in Section 4.



Figure 7: Distribution of travel times among new hires



The figure shows a histogram of the travel time to the new job among the sample of new hires out of unemployment. Travel time to new job is measured as the travel time between the centroids of the hired worker's zip code and the hiring firm's municipality. The figure only includes travel times of less than 200 minutes.

### 3.6 Summary statistics

The construction of the local unemployment and vacancy stocks in the previous section completes the data set used in the empirical analysis. Table 3 shows summary statistics.

Panel A shows the variables available at the zip code-year level. The number of observations here reflect that the analysis includes 567 zip codes in each of the years 2004-2006. The summary statistics for number of unemployed workers and vacancies within 60 minutes shows that the data contains very large variation in the size of local labor markets with both unemployed workers and vacancies varying by a factor of more than 80.

Panel B shows the continuous and binary variables available for the sample of hires. While most variables are available for the full sample, the indicator variable for whether the worker is reemployed in the same industry as his previous job is missing for about a third of the sample. This is due in part to workers who have not held a job before (and hence have no prior industry) and in part to missing industry data for some firms. In addition, the monthly wage variable is also missing for parts of the sample due to the trimming described in Section

3.1. When presenting empirical results later, the total sample size will thus change somewhat depending on the variables included in the regression.

Looking at the summary statistics for the monthly wage variable, it is also worth noting than even after the trimming described in Section 3.1, the data includes some surprisingly low monthly wages (the listed monthly wage is in Danish kroner, which currently about  $\frac{1}{5}$  of a dollar). This likely reflects some part time jobs as well as jobs where the employment relationship is reported to continue even after the worker has effectively stopped putting in new hours.<sup>32</sup>

Panel C finally shows the categorical variables available for the sample of hires. Again, the variable indicating the industry of the new job is missing for about a third of the sample due to missing information for some firms. The hires in the data are seen to be spread across establishments in 268 different Danish municipalities.<sup>33</sup>

## 4 Empirical results

I now turn to an empirical examination of the effect of labor market size on job search outcomes. Using the spatial variation in labor market size across local labor markets, I compare large and small markets and look for evidence that labor market size matters for job search outcomes. As with most of the existing literature, the possibility that unobserved differences between local labor markets can drive differences in job search outcomes is a concern for the interpretation of the results. I attempt to address this using control variables in the regression considered below.

### 4.1 Preliminary exercise: job finding rates

While the main focus of this paper is the relationship between labor market size and match quality, I start by examining job finding rates. As noted in the introduction, a large literature on the estimation of so-called matching functions have found that job finding rates generally do not vary with labor market size, hence I first examine whether this is reflected in my data as well.

For comparison with the previous literature, I present results from OLS estimation of the following regression, which is typical in the matching function literature:

---

<sup>32</sup>Since the monthly wage is based on dividing total wages paid by the total registered days of employment in a year, employment relationships whose length is overstated will tend to create very low monthly wages for some hires.

<sup>33</sup>The reason this is slightly smaller than the total number of municipalities is that I am dropping hirings involving a few municipalities, see footnote 22.

Table 3: Summary statics

<i>Panel A: Zip code-year variables</i>					
	Obs.:	Mean:	S.d.:	Min:	Max:
Unemployed within 0-60 min.	1,701	19,589	14,747	693	58,557
Vacancies within 0-60 min.	1,701	1,888	1,480	57	5,881
Job finding rate	1,701	0.83	0.19	0.27	1.87
Population	1,701	9,125	12,420	169	94,533
Area (km <sup>2</sup> )	1,701	73.42	71.57	1.26	456.67
<i>Panel B: Hires, continuous and binary variables</i>					
	Obs.:	Mean:	S.d.:	Min:	Max:
Same industry as previous job	194,629	0.18	0.39	0	1
Travel time to job	311,260	26.95	34.08	1	353
Monthly starting wage (kr)	291,865	18,732	8,380	935	71,904
Spell duration $\geq 2$ yrs	311,260	0.44	0.50	0	1
Female	311,260	0.52	0.50	0	1
Experience	311,260	11.12	8.95	0	42.65
Schooling	311,260	12.29	2.57	6	20
<i>Panel C: Hires, categorical variables</i>					
	Obs.:	Categories:			
Industry	199,773	112			
Education	311,260	344			
Municipality of work	311,260	268			

The table shows summary statistics for the data set used in the paper. Panel A shows summary statistics for variables available at the zip code by year level for 567 zip codes and three years. The job finding rate is defined as the total number of hired workers from the zip code divided the daily average number of unemployed workers in the zip code. Panel B shows summary statistics for the continuous and binary variables available at the level of an individual hire. The indicator for same industry as previous job is equal to one if the hiring firm is in the same industry as the previous firm the worker worked for. Travel time to new job is measured as the travel time between the centroids of the hired worker's zip code and the hiring firm's municipality. The indicator for a spell duration greater than two years is equal to one if the hired worker does not become unemployed in the two calendar years following the hire. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Panel C shows the number of observations with available data as well as the total number of unique values for each of the discrete, non-binary variables in the data. The industry and education variables here measure the industry of the hiring firm and the type of degree for the worker's highest held degree.

$$y_{ct} = \alpha_1 \log(u_{ct}) + \alpha_2 \log(v_{ct}) + \kappa_t + z'_{ct}\gamma + \varepsilon_{ct} \quad (12)$$

The unit of observation is a zip code-year<sup>34</sup> and the outcome variable,  $y_{ct}$ , is the log of the job finding rate (total new hires divided by the average number of unemployed in the zip code). On the right hand side  $\log(u_{ct})$  and  $\log(v_{ct})$  are the log number of unemployed workers and vacancies within the local labor market (within 60 minutes of the zip code), the  $\kappa_t$ 's are a set of year fixed effects and  $z_{tc}$  is a vector of zip code controls. The fact that local labor markets overlap in the data implies that the error terms in the regression,  $\varepsilon_{ct}$ , are correlated across space, so here and throughout the paper I report spatial correlation robust standard errors for all estimates (Conley (1999)), which are robust to both spatial correlation and correlation across years.<sup>35</sup>

The coefficients  $\alpha_1$  and  $\alpha_2$  in (12) capture the effect of being in a labor market with many unemployed workers and vacancies, respectively. The focus of the present paper is the effect of increasing the total market size - that is, the effect of simultaneously increasing both the number of unemployed workers and vacancies. The coefficient of interest is therefore  $\alpha_1 + \alpha_2$ . I report the implied estimate of this along with its standard error.

Table 4 shows OLS estimates of (12). The first column shows the estimates when only year fixed effects are included as controls. In line with previous estimates of matching functions, the results suggest that the number of unemployed workers in the local labor market has a negative and significant effect on the job finding rate, while the number of vacancies has a positive and significant effect. The implied effect of market size,  $\hat{\alpha}_1 + \hat{\alpha}_2$  is shown at the bottom. It is estimated to be small, negative and not statistically significant. Column (2) and (3) adds zip-code level controls, first the log of the zip code population and then the log of the zip code's area. Adding these controls only affects estimates very little. Overall, the data used in the present paper thus confirms the finding in the previous literature; job finding rates do not appear to vary with the total size of the labor market.

---

<sup>34</sup>Note, while the data I use in this paper is very well suited for examining the relationship between match quality and the number of unemployed workers and vacancies in the local labor market, it is less well suited to examine job finding rates because aggregating the time dimension to yearly data introduces so-called aggregation bias. Petrongolo and Pissarides (2001) provide a longer discussion (although their theoretical analysis does not apply directly here because my unemployment and vacancies data is yearly averages rather than snapshots at a point in time).

<sup>35</sup>In constructing the reported standard errors I use a Bartlett-kernel with a bandwidth of 60 minutes. Since labor market size only varies at the zip code-level and the asymptotic results of Conley (1999) here require that the number of zip codes goes to infinity, I also apply a version of the standard degree of freedom correction used for clustered standard errors:  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.

Table 4: Regression results, job finding rate

<i>Outcome: Log job finding rate</i>			
	(1)	(2)	(3)
Log unemployed workers ( $\alpha_1$ )	-0.2700*** (0.0401)	-0.2649*** (0.0401)	-0.2306*** (0.0415)
Log vacancies ( $\alpha_2$ )	0.2536*** (0.0418)	0.2538*** (0.0410)	0.2476*** (0.0419)
Log zip code population		-0.0132** (0.0066)	-0.0314*** (0.0098)
Log zip code area			0.0484*** (0.0106)
<i>Controls:</i>			
Year FEs	Yes	Yes	Yes
Observations:	1,709	1,709	1,709
<i>Implications for market size:</i>			
Implied effect of market size ( $\alpha_1 + \alpha_2$ )	-0.0164 (0.0161)	-0.0111 (0.0159)	0.0171 (0.0165)

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

The table reports OLS regression estimates. In each column the observations are zip code-years. The outcome variable is the log of the job finding rate as measured by the total number of hired workers in the zip code divided by the average number of unemployed worker in the zip code during the year. Unemployed workers is measured by the total number of unemployed workers within 60 minutes of the zip code. Vacancies is measured by the total number of vacancies within 60 minutes of the zip code. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.

## 4.2 Evidence from worker and firm observables

I now turn to the main focus of the paper: the relationship between labor market size and match quality among new hires. I first examine two indicators of match quality that are both based on the observables of the worker and firm involved in each new hire.

The first indicator is whether workers find a job in the same industry as their previous job. As discussed briefly in the introduction, if workers are assumed to have industry specific human capital from their previous jobs, workers will be more productive if reemployed in the same industry and hence jobs in that industry will be a better match.<sup>36</sup>

The second indicator is the travel time from the worker's home to his new job. The use of this as an indicator of match quality builds on the simple idea that because commuting is costly, jobs that are close by will be a better match. In line with this idea, existing evidence from Denmark suggests that workers indeed prefer jobs with shorter commutes and thus that firms have to compensate workers for longer commutes by paying higher wages (Vejlin (2010), Mulalic et al. (2010)).

The empirical results will be based on a regression equation similar to (12) but now written at the level of a hire (indexed by  $i$ ):

$$y_{ict} = \alpha_1 \log(u_{ct}) + \alpha_2 \log(v_{ct}) + \kappa_t + z'_{ict}\gamma + \varepsilon_{ict} \quad (13)$$

The left hand side variable,  $y_{ict}$ , will be one of the two indicators of match quality discussed above. On the right hand side,  $\log(u_{ct})$ ,  $\log(v_{ct})$  and  $\kappa_t$  are as before, while the vector of controls,  $z_{itc}$ , is now at the level of the individual hire. As before, the effect of labor market size on  $y_{ict}$  is captured by  $\alpha_1 + \alpha_2$ , however, instead of estimating (13) and then computing the implied effect of market size from the estimates of  $\alpha_1$  and  $\alpha_2$ , it is convenient to rewrite (13) in a way that makes the effect of labor market size appear directly.<sup>37</sup>

$$y_{ict} = (\alpha_1 + \alpha_2) \underbrace{\log(u_{ct})}_{\text{Market size}} + \alpha_2 \underbrace{\log\left(\frac{v_{ct}}{u_{ct}}\right)}_{\text{Market tightness}} + \kappa_t + z'_{ict}\gamma + \varepsilon_{ict} \quad (14)$$

As (14) shows, when  $\log(v_{ct})$  is replaced by  $\log\left(\frac{v_{ct}}{u_{ct}}\right)$  in the regression, the effect of market size,  $\alpha_1 + \alpha_2$ , can be read off directly as the coefficient on  $\log(u_{ct})$ . The intuitive reason for

---

<sup>36</sup>By a similar token, it would be interesting to examine also whether workers are reemployed in the same occupation as their previous job. Unfortunately, the administrative data which I use only contains occupation information for a small, highly select subsample of workers.

<sup>37</sup>I could of course also have used this reparameterization for (12) above, however, regression equations like (12) are more standard in the matching function literature and hence I focus on (12) for comparability.

this is that since the log ratio of vacancies to unemployed workers (the log market tightness) is included as a control in (14), the coefficient on  $\log(u_{ct})$  can be interpreted as the effect of increasing the number of unemployed workers, *while keeping the ratio of vacancies and workers constant*, which is of course just the same as simultaneously increasing both the number of unemployed workers and vacancies. When examining the effect of market size on match quality in the rest of the paper, I therefore present estimates of (14) and refer to  $\log(u_{ct})$  as the log of market size and  $\log\left(\frac{v_{ct}}{u_{ct}}\right)$  as log of market tightness.

Table 5 shows estimates of (14) when using the indicator for reemployment in the same industry as the outcome variable. The first column shows the estimated effect of market size in the baseline regression where the only controls are the log market tightness and the year fixed effect. The estimated coefficient on market size is 0.0133 and highly significant effect, implying that workers in larger labor markets are more likely to find a new job in the same industry as their previous job. Column (2) adds a set of basic Mincer controls (schooling, experience, experience squared and gender) to assess whether differences in worker characteristics across areas are driving results. The estimated coefficient is virtually unchanged so this does not appear to be the case. Column (3) adds the log of the zip code population and the log of the zip code area to assess whether results are being driven by differences in local population density, which has been shown previously to predict the frequency of industry and occupation changes. Including these zip code level controls is seen to lower the estimated coefficient slightly but the estimated effect remains positive and highly significant. Finally, Column (4) adds the full set of industry and education fixed effects to address differences in the industrial composition across areas as well as differences in worker types not captured by the Mincer controls. This barely changes the coefficient.

In sum, unemployed workers in larger labor market are significantly more likely to find a job which is a good match in terms of previous industry experience. Taking Column (4) as the preferred specification, the estimated effect of market size here implies that doubling the market leads to a 0.5 percentage point increase in the probability of being reemployed in the same industry. Given that labor market size varies by much more than a factor of two in the data (more than a factor of 80 in fact), this suggests that labor market size has substantial effects on job search outcomes.

While the focus of the present paper is the effects of labor market size, Table 5 also shows results on the effects of labor market tightness. In the preferred specification in Column (4), for example, the estimated coefficient on labor market tightness is -0.0117 and is significant at the 1% level. In Section 4.7 at the end I briefly discuss all the results I find regarding labor

Table 5: Regression results, reemployed in same industry

<i>Outcome: Reemployed in same industry as previous job</i>				
	(1)	(2)	(3)	(4)
Log market size	0.0133*** (0.0035)	0.0129*** (0.0037)	0.0085*** (0.0025)	0.0067*** (0.0021)
Log market tightness	-0.0212*** (0.0078)	-0.0204*** (0.0075)	-0.0193*** (0.0069)	-0.0117*** (0.0041)
Experience		0.0050*** (0.0007)	0.0051*** (0.0008)	0.0031*** (0.0007)
Experience squared <sup>†</sup>		-0.0117*** (0.0026)	-0.0120*** (0.0028)	-0.0060*** (0.0021)
Schooling		0.0052*** (0.0004)	0.0049*** (0.0004)	
Female		-0.0135*** (0.0037)	-0.0133*** (0.0036)	0.0036 (0.0033)
Log zip code population			-0.0014 (0.0012)	0.0001 (0.0009)
Log zip code area			-0.0057** (0.0025)	-0.0038*** (0.0012)
<i>Controls:</i>				
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	194, 629	194, 629	194, 629	194, 629

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

†: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is an indicator for whether the hire involves the worker being reemployed in the industry of his most recent job. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.



market tightness. I therefore postpone any further discussion of labor market tightness until Section 4.7.

The other indicator of match quality that I consider is the travel time to the new job. Table 6 shows estimates of (14), when this is used as the outcome variable. The specifications considered in the different columns of the table are the same as before. Without controls, as well as with the set of Mincer controls, the estimated coefficient is about -0.17 and significant at the 1% level, implying that workers in larger labor markets find jobs with shorter commutes. When I include zip code area and population area controls in Column (3), the estimated coefficient drops to about -0.10 and becomes significant only at the 5% level. Adding the full set of industry and education fixed effects in Column (4) further decreases the magnitude of the estimate a slightly but does not affect the statistical significance.

The results in Table 6 thus show that unemployed workers in larger labor markets find jobs that are a better match based on their geographical location. Again taking Column (4) as the preferred specification, doubling the size of the labor market implies a 5.8% decrease in the travel time to the new job. Converting to a yearly figure, this is equivalent to each worker saving 13.6 hours per year (at the mean).

As opposed to the results on reemployment in the same industry, which are difficult to convert into a monetary value, I can use a standard estimate of how workers value reductions in travel time to convert the estimated time saving into a monetary value. The standard figure used to value travel time reductions in Denmark is 77 kroner,<sup>38</sup> or about \$13 per hour saved. Based on this number, doubling the size of the labor market results in a welfare increase of 1048 kroner or \$182 per worker per year, which is about 0.6% of the mean starting wage in the data. As I return to in Section 4.6, effects of this magnitude suggest that the effects of labor market size are very important in practice.

### 4.3 Evidence from wages and employment spell durations

The results in the previous section showed that workers in larger labor markets find jobs that are better matches based on observable characteristics of the worker and the firm. Motivated by the theoretical results in Section 2, I now examine whether better match qualities also translate into higher wages and longer employment spells. I start by estimating the regression equation (14) using either the log monthly starting wage or an indicator for whether the associated employment spell lasted at least two years as the outcome variable.

Table 7 and 8 shows the results. In both tables, all specifications show that larger labor

---

<sup>38</sup>This figure is based on *Transportøkonomiske Enhedspriser 2010*, Technical University of Denmark .

Table 6: Regression results, travel time to new job

<i>Outcome: Log travel time to new job</i>				
	(1)	(2)	(3)	(4)
Log market size	−0.1661*** (0.0400)	−0.1792*** (0.0408)	−0.0977** (0.0412)	−0.0862** (0.0370)
Log market tightness	−0.1203** (0.0612)	−0.1157* (0.0598)	−0.0928 (0.0592)	−0.1274** (0.0550)
Experience		0.0032 (0.0024)	−0.0001 (0.0028)	−0.0002 (0.0015)
Experience squared <sup>†</sup>		−0.0079 (0.0051)	−0.0007 (0.0060)	0.0007 (0.0035)
Schooling		0.0319*** (0.0050)	0.0343*** (0.0048)	
Female		−0.2027*** (0.0117)	−0.2140*** (0.0138)	−0.1393*** (0.0215)
Log zip code population			−0.0957** (0.0468)	−0.0921* (0.0477)
Log zip code area			0.0533** (0.0264)	0.0595** (0.0243)
<i>Controls:</i>				
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	311, 260	311, 260	311, 260	199, 773

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

†: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is the log of the travel time between the centroids of the hired worker's zip code and the hiring firm's municipality. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.

markets are associated with significantly higher wages and longer employment durations. Based on Column (4), the estimates imply that doubling the size of the labor market increases wages by 3.2% and increases the probability that a new employment spell lasts at least two years by 1.6 percentage points.

These results show that workers in larger labor markets find jobs which pay higher starting wages and which result in longer employment spells. Interpreting this as evidence that labor market size affects match quality, however, is highly problematic because workers in larger labor markets are more likely to find work at firms in urban areas. These firms are generally believed to be more productive and in turn might pay higher wages and have longer employment spells for reasons unrelated to the size of the labor market. Therefore it is possible that the results in Table 7 and 8 merely reflect spatial productivity differences.

#### 4.4 Accounting for spatial productivity differences

As discussed above, the previous sections results regarding wages and employment spell durations are likely to be at least partly driven by spatial productivity differences among firms. To estimate the effect of market size on wages and employment spell durations, I therefore need a way to disentangle market size effects from the effects of working in a certain location. To do this, I utilize the fact that I can add municipality of work fixed effects to the regression equation (13):

$$y_{ict} = (\alpha_1 + \alpha_2) \log(u_{ct}) + \alpha_2 \log\left(\frac{v_{ct}}{u_{ct}}\right) + \kappa_t + z'_{ict}\gamma + \sum_j \omega_j \cdot d_{ict}^j + \varepsilon_{ict} \quad (15)$$

Here  $\omega_j$  is the municipality fixed effect for working in municipality  $j$ , while  $d_{ict}^j$  is an indicator variable for whether the hiring firm is located in municipality  $j$ . The inclusion of municipality fixed effects in the regression is made possible by the unique geographical disaggregation of the data: Even conditioning on where the worker ends up working, there is still variation left in the size of the worker's local labor market because workers are known to live in different zip codes. In a more standard data set where worker and firm locations are only identified by a common location variable, there would be no variation left in labor market size once the location of the firm was controlled for and the labor market variables would drop out of the regression.

In estimating (15), the effect of labor market size is identified by comparing workers who work in the same municipality but are located in different local labor markets (different zip codes). Figure 8 shows a graphical illustration: Worker 1 and 2 both work in municipality C

Table 7: Regression results, starting wage on new job

<i>Outcome: Log monthly wage</i>				
	(1)	(2)	(3)	(4)
Log market size	0.0485*** (0.0071)	0.0435*** (0.0069)	0.0383*** (0.0058)	0.0451*** (0.0070)
Log market tightness	0.0152 (0.0132)	0.0200* (0.0120)	0.0206* (0.0113)	0.0097 (0.0124)
Experience		0.0265*** (0.0013)	0.0267*** (0.0014)	0.0231*** (0.0014)
Experience <sup>2</sup>		-0.0588*** (0.0030)	-0.0592*** (0.0030)	-0.0536*** (0.0033)
Schooling		0.0350*** (0.0021)	0.0348*** (0.0020)	
Female		-0.2081*** (0.0159)	-0.2077*** (0.0161)	-0.1141*** (0.0124)
Log zip code population			0.0006 (0.0022)	0.0025 (0.0022)
Log zip code area			-0.0058* (0.0032)	-0.0094** (0.0041)
<i>Controls:</i>				
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	291, 575	291, 575	291, 575	189, 087

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

†: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is log of the monthly wage in the first calendar year. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes.

Table 8: Regression results, employment spell duration

<i>Outcome: Employment spell lasts at least 2 years</i>				
	(1)	(2)	(3)	(4)
Log market size	0.0297*** (0.0031)	0.0216*** (0.0033)	0.0253*** (0.0045)	0.0230*** (0.0039)
Log market tightness	0.0833*** (0.0068)	0.0834*** (0.0058)	0.0832*** (0.0059)	0.0718*** (0.0056)
Experience		-0.0019* (0.0010)	-0.0020** (0.0010)	-0.0037*** (0.0006)
Experience squared <sup>†</sup>		0.0059*** (0.0023)	0.0062*** (0.0022)	0.0107*** (0.0017)
Schooling		0.0183*** (0.0011)	0.0184*** (0.0011)	
Female		-0.0714*** (0.001)	-0.0717*** (0.001)	-0.0495*** (0.001)
Log zip code population			-0.0008 (0.0021)	-0.0007 (0.0020)
Log zip code area			0.0040** (0.0019)	0.0011 (0.0017)
<i>Controls:</i>				
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	311, 260	311, 260	311, 260	199, 773

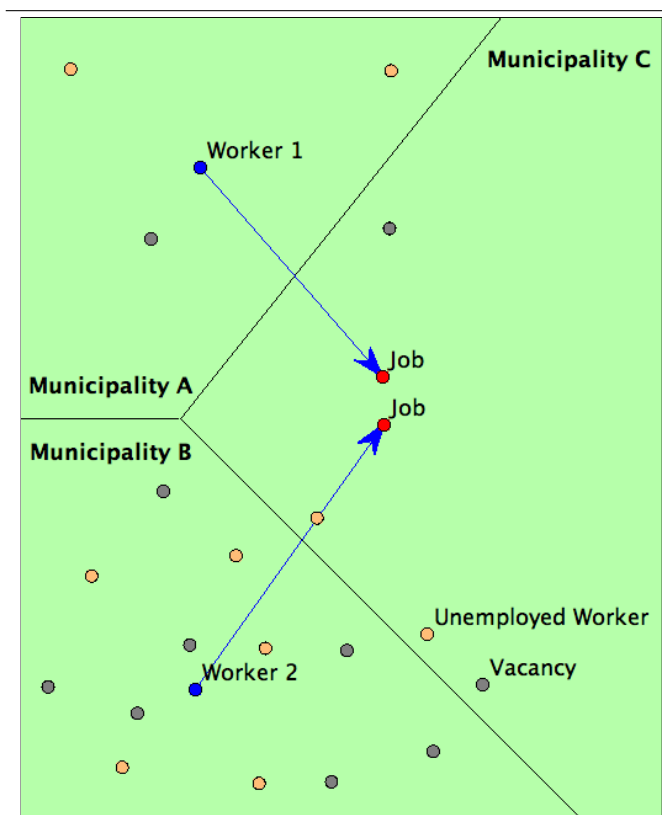
\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

†: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is an indicator for whether the hired worker does not become unemployed in the two calendar years following the hire. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.

but because they live in different in different locations, the size of their local labor market is different. Worker 1 lives in an area with a lot of vacancies and unemployed workers nearby and thus faces a large local labor market. Worker 2, on the other hand, lives in an area with few vacancies and unemployed workers nearby and faces a small local labor market. Since the two workers work at firms in the same municipality, a comparison across the two workers separates the effect of local labor market size from differences in municipal productivity levels.

Figure 8: Identification with municipality FEs



The figure shows the spatial configuration of two jobs and their two newly hired workers as well as a number of other unemployed workers and vacancies. The newly hired workers are blue dots and their jobs are red dots. Unemployed workers are represented by tan dots and vacancies by grey dots. Municipality borders are shown but zip code borders are omitted to avoid clutter. The two workers are assumed to live in separate zip codes however.

Tables 9 and 10 show estimates of (15) for wages and employment spell durations, respectively. The columns respond to the same specifications as the columns in Tables 5 through 8 but with municipality of work fixed effects added. Comparing the estimated effects of market size on wages and employment spell durations in Tables 9 and 10 with those in Tables 7 and 8

suggests that spatial productivity differences indeed do explain part of the association between labor market size and wages as well as employment spell durations. Including municipality of work fixed effects causes the estimated effect on wages to drop by about a fourth in Column (1) and by more than three fourths in the preferred specification in Column (4), where the estimate is also only marginally significant ( $p = 0.103$ ). For employment spell durations, the difference is less stark, but the estimates still decrease by between a fifth and a third. In both tables, however, the estimated effects of labor market size remains positive, sizeable and at least marginally significant. Even conditional on where they end up working, workers in larger labor markets still find jobs that pay higher wages and lead to longer employment spells. Taking Column (4) as the preferred specification, doubling the labor market size is estimated to increase wages by 0.6% and increase the probability of being reemployed in the same industry by 1.0%.

#### 4.5 Worker sorting

The results in the preceding section show that unemployed workers in larger labor markets find jobs that are better matches based on both industry and geographical location and that they also find jobs that pay higher wages and result in longer employment spells even after controlling for spatial productivity differences among firms. These results suggest that large labor markets improve job search outcomes by leading workers to find better matches. It is important, however, to consider other possible explanations for these findings.

An important alternative explanation is that different types of workers sort into different labor markets and that this is driving the observed differences in job search outcomes across areas. Although the results above are robust to controlling for worker observables, sorting may still take place based on unobservables. Combes et al. (2008), for example, provide evidence that workers with high unobserved abilities self-select into larger labor markets.

Two facts alleviate the concerns about worker sorting: The first is the fact that most of the estimates presented above change very little when worker characteristics are added as controls. If the results above were primarily driven by worker characteristics which are unobserved but most likely correlated with observable characteristics, then one would instead expect the inclusion of worker characteristics to cause large changes in estimates.<sup>39</sup>

The second alleviating fact is that the pattern of results presented above is not easily explained by high ability workers selecting into larger labor markets. While high ability workers could indeed be expected to earn higher wages and have longer employment spells, it is not

---

<sup>39</sup>Altonji et al. (2005) provide a formalization of this idea.

Table 9: Regression results, starting wage on new job, municipality FEs

<i>Outcome: Log monthly wage</i>				
	(1)	(2)	(3)	(4)
Log market size	0.0277*** (0.0102)	0.0262*** (0.0082)	0.0222*** (0.0077)	0.0081 (0.0050)
Log market tightness	-0.0331** (0.0149)	-0.0210** (0.0100)	-0.0205** (0.0096)	-0.0320*** (0.0082)
Experience		0.0262*** (0.0013)	0.0263*** (0.0013)	0.0229*** (0.0013)
Experience squared <sup>†</sup>		-0.0581*** (0.0028)	-0.0583*** (0.0028)	-0.0536*** (0.0032)
Schooling		0.0356*** (0.0022)	0.0355*** (0.0021)	
Female		-0.2020*** (0.0158)	-0.2010*** (0.0161)	-0.1110*** (0.0117)
Log zip code population			-0.0004 (0.0017)	-0.0003 (0.0010)
Log zip code area			-0.0066*** (0.0023)	-0.0013 (0.0024)
<i>Controls:</i>				
Municipality of work fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	291, 575	291, 575	291, 575	189, 087

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

†: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is log of the monthly wage in the first calendar year. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.



Table 10: Regression results, employment spell duration, municipality FEs

<i>Outcome: Employment spell lasts at least 2 years</i>				
	(1)	(2)	(3)	(4)
Log market size	0.0232*** (0.0047)	0.0162*** (0.0047)	0.0189*** (0.0043)	0.0147*** (0.0036)
Log market tightness	0.0632*** (0.0131)	0.0675*** (0.0112)	0.0675*** (0.0113)	0.0692*** (0.0131)
Experience		-0.0021** (0.0010)	-0.0021** (0.0009)	-0.0037*** (0.0006)
Experience squared <sup>†</sup>		0.0062*** (0.0022)	0.0064*** (0.0022)	0.0107*** (0.0016)
Schooling		0.0182*** (0.0011)	0.0182*** (0.0011)	
Female		-0.0698*** (0.0096)	-0.0700*** (0.0095)	-0.0490*** (0.0092)
Log zip code population			-0.0012' (0.0020)	-0.0006 (0.0016)
Log zip code area			0.0032 (0.0028)	0.0014 (0.0023)
<i>Controls:</i>				
Municipality of work fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes
Education fixed effects	No	No	No	Yes
Observations:	311, 260	311, 260	311, 260	199, 773

\*: Significant at 10%, \*\*: Significant at 5%, \*\*\*: Significant at 1%

<sup>†</sup>: Experience squared is scaled down by 100 to keep coefficients of similar magnitude.

The table reports OLS regression estimates. In each column the sample consists of all new hires out of unemployment without missing information on the outcome or control variables. The outcome variable is an indicator for whether the hired worker does not become unemployed in the two calendar years following the hire. Market size is measured by the total number of unemployed workers within 60 minutes of the hired worker's zip code. Market tightness is measured by the ratio of vacancies to unemployed workers within 60 minutes of the hired worker's zip code. Experience is measured as years of previous employment. Schooling is measured by the years required to complete the worker's highest held degree. Industry fixed effects is based on the industry of the firm using a classification of 112 industries. Education fixed effects are based on the exact type of the worker's highest held degree using a classification with 344 different types of degrees. The table reports spatial correlation robust standard errors (Conley (1999)), which are robust to both spatial correlation and correlation across years. The reported standard errors are constructed using a Bartlett-kernel with a bandwidth of 60 minutes as well as a conservative degrees of freedom correction factor of  $\frac{M}{M-1} \cdot \frac{N}{N-K}$ , where  $M$  is the number of zip codes,  $N$  is the total number of observation and  $K$  is the number of estimated coefficients.

obvious that in a labor market where all workers have higher ability, they would *all* be more likely to find a job in their previous industry and close to where they live.

Despite these alleviating facts, the possibility that worker sorting on unobservables contributes to the observed differences across labor markets remains a caveat when interpreting this paper's results.

## 4.6 The practical importance of labor market size

I now examine the economic significance of the estimated effects of labor market size. I do this in two simple ways: First, I examine the role of labor market size in shaping spatial wage differences. Second, I perform a simple back-of-the-envelope calculation of the wage benefits associated with one specific transport infrastructure project.

To examine the role of labor market size in explaining spatial differences in wages, I revisit the wage regression with municipality fixed effects from Table 9 and interpret the estimated municipality fixed effects as estimates of the municipal wage premia. I then estimate the municipality wage premia both with and without controlling for labor market size, and compare the variance of the estimated fixed effects to measure of the fraction of the variation in spatial wage premia that can be explained by labor market size.

I focus on the preferred specification with all controls in Column (4). The variance of the estimated fixed effects in this column is 0.0054 when the measure of market size,  $\log(u_{ct})$ , is dropped from the regression. With market size included this variance drops to 0.0051, suggesting that labor market size can explain about 6.6% of the spatial variation in wage premia. Thus labor market size appears to play a sizable role in shaping spatial differences in wages.

Next, to get a sense of how important labor market size effects are for the benefits of transport infrastructure projects, I perform a simple calculation of the wage effects of one particular transport infrastructure project: the San Francisco-Oakland Bay Bridge in California. The bridge was completed in 1936 and connects the labor markets of San Francisco and Oakland, which would otherwise be separated by water. Below I use the estimated effect of labor market size on wages to compute the total loss in wage earnings that would occur if the Bay Bridge was removed and the labor markets of San Francisco and Oakland were separated.

I start by computing the total wage earnings in the two labor markets. The population of the San Francisco and Oakland Census County Divisions were 802,235 and 411,480, respectively, at the time of the 2010 census. Based on the overall employment-to-population ratio for California and the average yearly earnings for employees in San Francisco-Oakland-Fremont MSA, this

implies that total yearly wage earnings in San Francisco and Oakland are 25.0 billion and 12.8 billion, respectively.<sup>40</sup>

Under the assumption that the removal of the bridge does not change labor market tightness, unemployment rates or labor force participation, the drop in labor market size that would occur upon removing the bridge can be found by comparing the total population of San Francisco and Oakland to the population of each of them separately. Doing so shows that the removal of the Bay Bridge would decrease the labor market size for workers in San Francisco by 34%<sup>41</sup> and decrease the labor market size for workers in Oakland by 66%.<sup>42</sup> Using the estimated effect of labor market size on wages from Column (4) of Table 9, this implies that removing the bridge would lower wages in San Francisco by 0.3% and in Oakland by 0.9%. Based on the yearly earnings totals this implies a loss in yearly earnings of \$83 million in San Francisco and \$112 million in Oakland. Thus the total yearly loss in wage earnings caused by the removal of the Bay Bridge would be \$194 million per year.

This loss can be benchmarked against the cost of the constructing the bridge. The original cost of constructing the bridge was \$77 million in November 1936, which inflated by the CPI equals \$1.3 billion in August 2012. Compared to the estimated wage increase of \$194 million per year, this suggests that the positive effect of labor market size on wages alone would cover 15.4% of the total cost of reconstructing the bridge already in the first year. Assuming that the bridge stays in place forever and has no maintenance costs, this also implies an internal rate of return of 15.4% based on the wage effects alone.

This back-of-the-envelope calculation is of course much too simple to be used for practical policy evaluation and may well overstate the wage returns of the Bay Area Bridge due to rising construction costs since 1936 and the assumption of no maintenance costs after initial construction.<sup>43</sup> It does however quite clearly illustrate that the estimated effects of labor market size are economically significant and warrants attention when formulating policy.

---

<sup>40</sup>The employment-to-population rate in California was 0.56 and the average yearly earnings for an employee in San Francisco-Oakland-Freemont MSA was \$55,470 in August 2012 according to the BLS.

<sup>41</sup>For San Francisco:  $\frac{802,235}{802,235+411,480} - 1 \approx -0.34$ . For Oakland  $\frac{411,480}{802,235+411,480} - 1 \approx -0.66$ .

<sup>42</sup>One might argue that this calculation overstates the size of the decrease in labor market size because the relevant labor market for workers on either side of the bridge also include adjacent areas (such as Berkeley), which can be reached regardless of the existence bridge. To account for this I can redo the calculation assuming that these adjacent areas increase labor markets by 20% regardless of the bridge. Under this assumption the total yearly effect on wages drops to \$146 million, corresponding to 11.2% of the cost of the bridge being covered in the first year and an internal rate of return of 11.2%.

<sup>43</sup>After a long string of problems and massive cost-overruns, an ongoing upgrade/repair effort on the Bay Area Bridge, which replaces the bridge's entire eastern span, is set to be finished later this year at a total cost of over \$ 6 billion (compared to an initial estimate of around \$ 1 billion). Setting the total cost of the bridge in the calculation above to \$ 7 billion drops the internal rate of return on the wage effects of the bridge to 2.7 %.

## 4.7 Labor market tightness and match quality

The focus of this paper is the effects of labor market size, however, since the regression results presented also include estimates of the effect of labor market tightness, I now briefly discuss these as well.

From a theoretical standpoint, the effect of labor market tightness on match quality is highly ambiguous. On the one hand, workers will have a lot of choice in who they match with when the labor market is tight. On the other hand firms have very little choice in tight markets. As a result, theoretical predictions on the effects of labor market tightness on match quality depends on the specifics of the job search process.

Looking first at the empirical results in the specification with all controls but no municipality fixed effects (Column (4) of Tables 5 through 8) reveals the following pattern: labor market tightness has a significant negative effect on reemployment in the same industry, has a highly significant negative effect on travel time to the new job, has no significant effect on wages and has a significant positive effect on employment spell durations. Besides the negative effect on reemployment in the same industry, these results seem to suggest that tight labor markets improve match quality, at least along certain dimensions.

Moving to the specifications where municipality of work fixed effects are added to the wage and employment duration regression in Tables 8 and 9, however, a fairly puzzling finding appears as labor market tightness is estimated to have a significant negative effect on wages. This finding is not consistent with match quality being higher in tighter markets and is particularly puzzling because tighter market should improve workers' bargaining position which should increase wages if they are set via a standard bargaining procedure. It is possible that a richer model of wage determination can explain this result, however, I leave further analysis of this for future work.

## 5 Conclusion

While much existing research and many popular policy initiatives build on the premise that larger labor markets lead to better job search outcomes, very little empirical evidence on this effect actually exists. On the contrary, a large literature relating job finding rates to the number of unemployed workers and vacancies has suggested that the job search process is unaffected by overall labor market size.

In response, the present paper examines the possibility that labor market size does have positive effects on the job search process but that this manifests only in workers finding jobs

for which they are a better match and not in the rate at which job finding occurs. I first show the possibility of this theoretically and next use a unique new data set to examine the relationship between local labor market size and the quality of the worker-job matches that occur across local labor markets in Denmark.

I find that workers in large labor markets find jobs that are better matches for them as measured by both previous industry experience and geographical location. In line with my theoretical analysis, workers in larger labor markets also earn higher starting wages in their new jobs and have longer employment spells, a finding that holds even conditional on where the workers find jobs. Finally I show that labor market size plays a sizable role in shaping spatial differences in wages and that the estimated effects of labor market size can imply large rates of return on transport infrastructure projects. In sum, the paper provides evidence that larger labor markets do in fact improve job search outcomes and that these effects are of a magnitude that warrants attention both in research and policy circles.

The paper's results suggest several avenues for future research. In relation to the theoretical job search literature, the results suggest that it may be fruitful to devote more attention to models of job search in which match qualities improve in larger labor markets. In addition, while the results presented here suggest that policies which increase the effective size of labor markets can have important benefits, a proper evaluation of specific policy initiatives would require a detailed model of job search in local labor markets. In ongoing work, I am therefore developing and estimating a spatial job search model, which can evaluate specific policy initiatives, while accounting for general equilibrium effects and geographical mobility. Finally, it seems very promising to also relate the present paper's results to firm location choice, as well as to on-the-job search and differences in workers' career trajectories. The data set used in the present paper are likely to be useful for such endeavors as well.

## References

- Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1), 151–184.
- Anderson, P. M. and S. M. Burgess (2000). Empirical matching functions: Estimation and interpretation using disaggregate data. *Review of Economics and Statistics* 82(1), 112–132.
- Andersson, F., S. Burgess, and J. I. Lane (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics* 61(1), 112–128.
- Bennet, R. J. and R. R. Pinto (1994). The hiring function in local labour markets in Britain. *Environment and Planning* 26(12), 1957–1974.
- Berman, E. (1997). Help wanted, job needed: Estimates of a matching function from employment service data. *Journal of Labor Economics* 15(1), 251–291.
- Blanchard, O. J. and P. A. Diamond. (1990). *The Aggregate Matching Function*. MIT Press.
- Bleakley, H. and J. Lin (2012). Thick-market effects and churning in the labor market: Evidence from US cities. *Journal of Urban Economics* 72(2-3), 87–103.
- Burda, M. C. (1993). Unemployment, labour markets and structural change in eastern Europe. *Economic Policy* 16, 101–137.
- Burda, M. C. and C. Wyplosz. (1994). Gross worker and job flows in Europe. *European Economic Review* 38(6), 1287–1315.
- Burgess, S. and S. Profit (2001). Externalities in the matching of workers and firms in Britain. *Labour Economics* 8(3), 313–333.
- Coles, M. G. and E. Smith (1996). Cross-section estimation of the matching function: Evidence from England and Wales. *Economica* 63(252), 589–598.
- Combes, P. P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2), 723–742.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics* 92(1), 1–45.

- Costa, D. L. and M. E. Kahn (2000). Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990. *The Quarterly Journal of Economics* 115(4), 1287–1315.
- Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics* 4, 2063–2117.
- Fahr, R., S. U. (2006). Spatial mobility and competition for jobs: some theory and evidence for western germany. *Regional Science and Urban Economics* 36(803–825).
- Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional Wage Dispersion in Search Models: A Quantitative Assessment. *American Economic Review* 101(7), 2873–2898.
- Hynninen, S.-M. (2005). Matching across space: evidence from finland. *LABOUR* 19(4), 749–765.
- Jovanovic, B. (1979). Job matching and the theory of turnover. *The Journal of Political Economy* 87(5), 972–990.
- Layard, R., S. J. Nickell, and R. A. Jackman (1991). *Unemployment, Macroeconomic Performance and the Labour Market*. Oxford University Press.
- Lottmann, F. (2012). Spatial dependencies in German matching functions. *Regional Science and Urban Economics* 42(1-2), 27–41.
- Manning, A. and B. Petrongolo (2011). How Local Are Labor Markets?: Evidence from a Spatial Job Search Model. *CEPR Discussion Papers*.
- Mion, G. and P. Naticchioni (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue canadienne d'économie* 42(1), 28–55.
- Moretti, E. (2011). Local labor markets. *Handbook of labor economics* 4, 1237–1313.
- Mortensen, D. T. and C. A. Pissarides (1999). New developments in models of search in the labor market. *Handbook of labor economics* 3, 2567–2627.
- Mulalic, I., J. Van Ommeren, and N. Pilegaard (2010). Wages and Commuting: Quasi-Natural Experiments' Evidence from Firms that relocate. *Tinbergen Institute Discussion Papers*.
- Munich, D., J. Svejnar, and K. Terrell (1998). Worker-firm matching and unemployment in transition to a market economy: (why) were the czechs more successful than others? *The William Davidson Institute Working Paper* 7.

- Papageorgiou, T. (2010). Worker Sorting and Agglomeration Economies. *Working Paper*.
- Petrongolo, B. and C. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39(2), 390–431.
- Petrongolo, B. and C. A. Pissarides (2006). Scale Effects in Markets with Search. *The Economic Journal* 116(508), 21–44.
- Pissarides, C. A. (1986). Unemployment and vacancies in Britain. *Economic Policy* 3, 676–690.
- Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics* 4, 2120–2171.
- Sahin, A., J. Song, G. Topa, and G. Violante (2012). Mismatch Unemployment. *Federal Reserve Bank of New York Staff Report* 566.
- Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95(1), 25–49.
- van Ours (1991). The efficiency of the Dutch labour market in matching unemployment and vacancies. *De Economist* 139(3), 358–78.
- Vejlin, R. (2010). Residential Location, Job Location, and Wages: Theory and Empirics. *Economics Working Papers*.
- Warren, R. S. (1996). Returns to scale in a matching model of the labour market. *Economic Letters* 50(1), 135–142.
- Wheeler, C. H. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics* 38(2), 101–118.
- Yashiv, E. (2000). The determinants of equilibrium unemployment. *American Economic Review* 90(5), 1297–1322.



## List of Tables

1	Choice of parameters and distributions for numerical analysis . . . . .	13
2	The effect of a 50% increase in market size . . . . .	15
3	Summary statics . . . . .	27
4	Regression results, job finding rate . . . . .	29
5	Regression results, reemployed in same industry . . . . .	32
6	Regression results, travel time to new job . . . . .	34
7	Regression results, starting wage on new job . . . . .	36
8	Regression results, employment spell duration . . . . .	37
9	Regression results, starting wage on new job, municipality FEs . . . . .	40
10	Regression results, employment spell duration, municipality FEs . . . . .	41

## List of Figures

1	Numerical results on the effect of market size . . . . .	13
2	Comparing survey-based and jobnet.dk vacancy measures . . . . .	19
3	Total jobnet.dk vacancies over time . . . . .	20
4	Danish zip codes and Municipalities . . . . .	21
5	Centroids of the Danish zip codes and municipalities . . . . .	22
6	Measuring the distances between centroids . . . . .	23
7	Distribution of travel times among new hires . . . . .	25
8	Identification with municipality FEs . . . . .	38
9	Measuring local labor market size I . . . . .	57
10	Measuring local labor market size II . . . . .	58

# Appendices

## A Derivation of model equations

Given the model and notation from Section 2, we can let  $W(\theta)$  denote the value to a worker of being in a job with match quality  $\theta$  and write down a standard asset pricing equation for the flow value of unemployment:

$$\rho U = b + \lambda_u(u, v) \int_0^\infty \max \{W(\theta) - U, 0\} dF(\theta; u, v) \quad (16)$$

The right hand side of this equation reflects that an unemployed worker earns a flow utility of  $b$  at each point in time, and at rate  $\lambda_u(u, v)$  he meets a vacant job. When he meets a vacant job, he learns his match quality with the job  $\theta \sim F$  and then optimally chooses to either accept the job and gain the difference between value of employment and unemployment  $W(\theta) - U$  or reject the job causing no change.

Letting  $J(\theta)$  be the value to the firm of a filled job with match quality  $\theta$  we can write a similar equation for the present value of a vacant job, which has just posted a vacancy notice:

$$V_0 = \int_0^\phi -a e^{-\rho t} dt + e^{-\rho \phi} \int_0^\infty \max \{J(\theta), V_0\} dF(\theta; u, v) \quad (17)$$

The right hand side of this equation consists of two terms, where the first one reflects the firm with the vacant job must first wait a period of length  $\phi$  while incurring the flow cost  $a$ . The second reflects that after the waiting period, the firm will receive applications from a number of workers and choose to meet with the best one, which again implies meeting a worker with match quality,  $\theta \sim F$ . Upon meeting the worker the firm will then optimally choose to either hire the worker, the value of which is  $J(\theta)$ , or keep the job vacant and post another vacancy notice, the value of which is  $V_0$ .

We can also write an asset pricing equation for the value of a filled job with match quality  $\theta$  to the involved worker-firm pair. We will denote this by  $Y(\theta)$ :

$$\rho Y(\theta) = \theta + \delta_0 (-Y(\theta) + U + V_0) + \delta_1 \int_0^\infty \max \{-c, -Y(\theta) + U + V_0\} dH(c) \quad (18)$$

The right hand side here reflects that at each point in time, the filled job produces a net output of  $\theta$  and is hit by exogenous separation shocks at rate  $\delta_0$  and a negative productivity shock at rate  $\delta_1$ . When hit by an exogenous separation shock, the worker and job separate

causing causing a loss of the total value  $Y(\theta)$  but also a total gain of  $U + V_0$  for the pair because the worker enters unemployment and the job becomes vacant. When hit by the exogenous separation shock, the worker-firm pair observes the cost of staying matched  $c \sim H$  and then optimally chooses between paying the cost  $c$  or separating, which again implies a loss of  $Y(\theta) - U - V_0$ .

Now by definition, the surplus from a type  $\theta$  match is  $S(\theta) = Y(\theta) - U - V_0$ , which we can use to rewrite (18) as:

$$\rho(S(\theta) + U + V_0) = \theta + \delta_0(-S(\theta)) + \delta_1 \int_0^\infty \max\{-c, -S(\theta)\} dH(c) \quad (19)$$

Moving  $\rho U$  and  $\rho V_0$  to the other side the yields equation (7) from the main text, which is the asset pricing equation for the surplus of a match.

Nash-bargaining implies that the surplus from a match is split so that the worker and job are compensated for their outside option of remaining unemployed or vacant and in addition receives a fixed share of the generated surplus:

$$W(\theta) = \beta S(\theta) + U \quad (20)$$

$$J(\theta) = (1 - \beta)S(\theta) + V_0 \quad (21)$$

Rewriting this and plugging in for  $W(\theta) - U$  and  $J(\theta) - V_0$  in (16) and (17) yields:

$$\rho U = b + \lambda_u(u, v) \int_0^\infty \max\{\beta S(\theta), 0\} dF(\theta; u, v) \quad (22)$$

$$V_0 = \int_0^\phi -a e^{-\rho t} dt + e^{-\rho \phi} \int_0^\infty \max\{(1 - \beta)S(\theta) + V_0, V_0\} dF(\theta; u, v) \quad (23)$$

Nash-bargaining also implies that a worker and job who meet choose to match only if the implied surplus is positive. Since  $S$  is increasing, this defines a reservation output  $\bar{\theta}$  so that a worker and job with match quality  $\theta$  chooses to matches if  $\theta > \bar{\theta}$ :

$$S(\bar{\theta}) = 0 \quad (24)$$

Now (24) can be used to eliminate the max operator from under the integral in (22) and (23), yielding equations (5) and (6) from the main text. Combining (24) with (19) also yields equation (8) from the main text.

To figure out the wages paid to a type  $\theta$  match, we start by considering a matched worker-firm pair who is hit by a negative productivity shock and chooses to pay the cost  $c$  to stay

together. If we let  $r(c)$  denote the part of the cost paid by the worker, Nash-bargaining implies:

$$W(\theta) - r(c) = \beta(S(\theta) - c) + U \quad (25)$$

Combining this with (20) shows that  $r(c) = \beta c$ , that is the worker pays a share  $\beta$  of the cost of staying matched. We can then write out an asset pricing equation for the value of employment:

$$\rho W(\theta) = w(\theta) + \delta_0(-W(\theta)) + \delta_1 \int_0^\infty \max\{-\beta c, -W(\theta)\} dH(c) \quad (26)$$

This equation of course just reflects that an employed worker earns a wage of  $w(\theta)$  at each point in time and then may experience exogenous separations or negative productivity shocks, which either cause him to pay part of the cost of staying matched or to separate.

Combining (26), (7) and (20) we can solve for  $w(\theta)$  to get equation (9) from the main text, which is a standard wage equation and shows that wages are a convex combination of the flow net output from the match and the flow value of being unemployed, with the weight depending on workers' bargaining power parameter.

Next we derive steady state conditions for employment and vacancies. As is clear from (19), a worker-firm pair that is hit by a negative productivity shock chooses not to separate if the associated cost satisfies  $c < S(\theta)$  that is they choose to stay matched if the value of the match surplus is greater than the cost of staying matched. Since  $c \sim H$  the probability of separation following a negative productivity shock is thus  $1 - H(S(\theta))$ . Taking exogenous separations into account as well, the total separation rate for a matched worker-firm pair with match quality  $\theta$  is then:

$$\delta(\theta) = \delta_0 + \delta_1 (1 - H(S(\theta))) \quad (27)$$

Since  $S(\theta)$  is increasing in  $\theta$ , the separation rate for a match is seen to be decreasing in match quality,  $\theta$ . This reflects that high quality matches are more valuable and hence that worker-firm pairs with a high match quality are more likely to choose to pay the cost and stay matched when hit by a negative productivity shock.

Because separation rates vary with match quality, the usual simple steady state equation for employment or unemployment does not apply. In particular, it is necessary to explicitly keep track of the distribution of match qualities among the employed,  $K$ .

To do this, we start by noting that the flow of workers into employment with match quality less than  $\theta$  is simply the total number of unemployed workers who meet a vacancy  $u\lambda_u(u, v)$

times the share of these who get a match quality draw above  $\bar{\theta}$  but below  $\theta$ :

$$u\lambda_u(u, v) (F(\theta; u, v) - F(\bar{\theta}; u, v)) \quad (28)$$

By integrating the separation rate from (27) over the distribution of match qualities among the employed and multiplying by the total number of employed workers,  $l - u$ , we can write the flow out of employment at a match quality of  $\theta$  or less as:

$$(l - u) \int_{\bar{\theta}}^{\theta} \delta(\theta') dK(\theta') \quad (29)$$

In equilibrium, the flow in and out of employment at less than  $\theta$  must be equal to each other for all levels of  $\theta$ , hence setting (28) and (29) equal we get:

$$u\lambda_u(u, v) (F(\theta; u, v) - F(\bar{\theta}; u, v)) = (l - u) \int_{\bar{\theta}}^{\theta} \delta(\theta') dK(\theta') \quad (30)$$

Rewriting and using (27) this then yields equation (10) from the main text, which pins down steady state employment at all levels of match quality and also steady state unemployment.

Finally we need to determine the steady state number of vacancies. Subtracting off unemployed workers,  $u$ , from the total number of workers,  $l$ , and subtracting off vacant jobs,  $v$  from the total number of jobs,  $n$ , yields the number of employed workers and the number of filled jobs, respectively. However, these two quantities must clearly be equal, hence we get the equation:

$$l - u = n - v$$

Rewriting then finally yields the steady state equation for vacancies, (11).

## B Numerical solution procedure

The main issue in solving the equilibrium job search model is that the equilibrium equations (2) and (4)-(11) include the two unknown functions  $S$  and  $K$  and the corresponding two functional equations (7) and (10).

To provide a numerical solution to the model, start by approximating the  $G$  distribution by a truncated version that has support on  $[0, \psi]$  so that the state space of match qualities is bounded. Next, note that since the assumed distribution for  $G$  is absolutely continuous, the same must be true for the  $F$  and  $K$  distributions. Letting lower case letters denote probability

density functions, we can then take derivatives on both sides of (10) and rearrange to get:

$$k(\theta) = \frac{u\lambda_u(u, v)f(\theta; u, v)}{(l - u)(\delta_0 + \delta_1 H(S(\theta)))} \quad (31)$$

Since  $K$  is a distribution function with support on  $[\bar{\theta}, \psi]$  we can integrate both sides of this equation over the support to get:

$$1 = \int_{\bar{\theta}}^{\psi} \frac{u\lambda_u(u, v)f(\theta; u, v)}{(l - u)(\delta_0 + \delta_1 H(S(\theta)))} d\theta \quad (32)$$

By replacing (10) by (32) in the set of equations characterizing the equilibrium it is no longer necessary to solve for the  $K$ -function when solving the model.

Next, we will replace the  $S$  function in the equilibrium equations by a new function  $S^*(x)$ , defined by  $S^*(x) = S((x - 1)(\psi - \bar{\theta}) + \bar{\theta})$ . Thus defined,  $S^*(x)$  is simply a version of  $S$  with domain  $[-1, 1]$  instead of  $[\bar{\theta}, \psi]$ . Since  $S(\theta) = S^*\left(\frac{\theta - \bar{\theta}}{\psi - \bar{\theta}}\right)$ , it is straightforward to replace  $S$  by  $S^*$  in all the equations characterizing the model equilibrium. Writing the functional equation, (7), in terms of  $S^*$  we get:

$$\rho S^*(x) = (x - 1)(\psi - \bar{\theta}) + \bar{\theta} - \rho U - \rho V_0 + \delta_0 (-S^*(x)) + \delta_1 \int_0^{\infty} \max\{-c, -S^*(x)\} dH(c) \quad (33)$$

The presence of the unknown  $S^*$ -function and the functional equation (33) can now be dealt with using a projection method. In particular, consider approximating the  $S^*(x)$  by a  $q$ th order polynomial  $\hat{S}^*$  with coefficient vector  $\mu = (\mu_0, \mu_1, \dots, \mu_q)$ .<sup>44</sup> To determine the coefficients,  $\mu$ , we will use Chebyshev collocation. In other words, we will pin down  $\mu$  by ensuring that (33) is satisfied at each of the  $q$  chebyshev nodes on  $[-1, 1]$ . Letting  $x_0, x_1, \dots, x_q$  denote the  $q + 1$  chebyshev nodes, this implies that the following set of equations replace (33) :

$$\begin{aligned} \rho S^*(x) &= (x - 1)(\psi - \bar{\theta}) + \bar{\theta} - \rho U - \rho V_0 + \delta_0 (-S^*(x)) + \delta_1 \int_0^{\infty} \max\{-c, -S^*(x)\} dH(c) \\ \text{for } x &= x_0, x_1, \dots, x_q \end{aligned} \quad (34)$$

With this change, the set of equations characterizing the equilibrium has been reduced to a finite set of non-linear equations in scalar unknowns which can be solved using standard methods for non-linear equations (and numerical integration).

---

<sup>44</sup>In the numerical work in the paper I choose  $q = 15$ .

## C The theoretical effects of labor market size, details

This section provides a longer discussion of how labor market size affect outcomes other than the job finding rate in the two-sided search model considered in the paper.

Starting with match quality, the average match quality among new hires in the model is simply found by integrating over the match quality distribution conditional on matching ( $\theta \geq \bar{\theta}$ ):

$$\int_{\bar{\theta}}^{\infty} \theta dF(\theta; u, v | \theta \geq \bar{\theta}) \quad (35)$$

Correspondingly, the effect of market size on match qualities is straightforward: since the unconditional match quality distribution,  $F$ , improves with market size, so does the truncated one, implying higher average match qualities for any given level of  $\bar{\theta}$ . Increases in the reservation output  $\bar{\theta}$  of course reinforces this effect by shifting the truncation point upward.

The average wage among new hires is simply:

$$\int_{\bar{\theta}}^{\infty} w(\theta) dF(\theta; u, v | \theta > \bar{\theta}) \quad (36)$$

Since the wage is an increasing function of match quality ( $w(\theta) = \beta\theta + (1 - \beta)\rho U$ ), one effect of larger labor markets on wages is that they increase the average match quality, which leads to higher wages. There is, however, another effect because changes in the size of the labor market also change the value of unemployment  $U$ , which affects wages as well. Under the parameters considered in the paper, larger labor markets lead to a higher value of unemployment, which implies another positive effect of labor market size on wages. This explains the relatively large positive effect of labor market size on wages.

Finally, the average separation rate among new hires is:

$$\int_{\bar{\theta}}^{\infty} (\delta_0 + \delta_1(1 - H(S(\theta)))) dF(\theta; u, v | \theta > \bar{\theta}) \quad (37)$$

For a fixed  $S$ -function, it is easy to see that the integrand above is decreasing in  $\theta$ , implying that larger labor markets lead to a lower average separation rate by improving match qualities. Since the surplus from a type  $\theta$  match,  $S(\theta)$ , generally also changes in response to market size, there is, however, also another effect. Because larger labor markets have a higher value of unemployment (and vacant jobs), this lowers the surplus from a match, leading to higher separation rates. The intuition is that better match qualities among new hires make it more

valuable for workers to be unemployed and jobs to be vacant. For the average separation rate among new hires, there are thus two offsetting effects: Higher match qualities among the new hires that form lead to lower separation rates, however, increases in the ease of finding good matches makes workers and firms more willing to separate. Under the parameters considered in the paper, these two effects are seen to net out to a moderate decrease in the average separation rate among new hires.

## D Measuring new hires out of unemployment, details

The data on new hires out of unemployment are constructed from data on employment spells and unemployment which do not have within-year dating of events. This is done by assuming that an employment spell in some year is a new hire out of unemployment in that year only if both of the following conditions are satisfied:

1. The involved worker-firm pair did not have a previous employment relationship in the current or previous year.<sup>45</sup> If a worker has multiple separate employment relationships with the same firm in a given year, the longest relationship is assumed to be the first one.
2. The involved worker *could* have been unemployed for at least three weeks prior to the employment relationships start, given the share of the year the worker has been unemployed and the number and length of the workers' other employment spells in the given year.<sup>46</sup>

## E Measuring the size of local labor markets, details

This section describes how I determine the number of vacancies and unemployed workers that are within  $t$  minutes of a zip codes centroid based on the number of unemployed workers and vacancies in each municipality as well as the travel times between the zip code centroid and municipality centroids.

To figure out the number of unemployed workers and vacancies from municipality A that are within  $t$  minutes of some zip code, I do the following: Let B denote the municipality that has

---

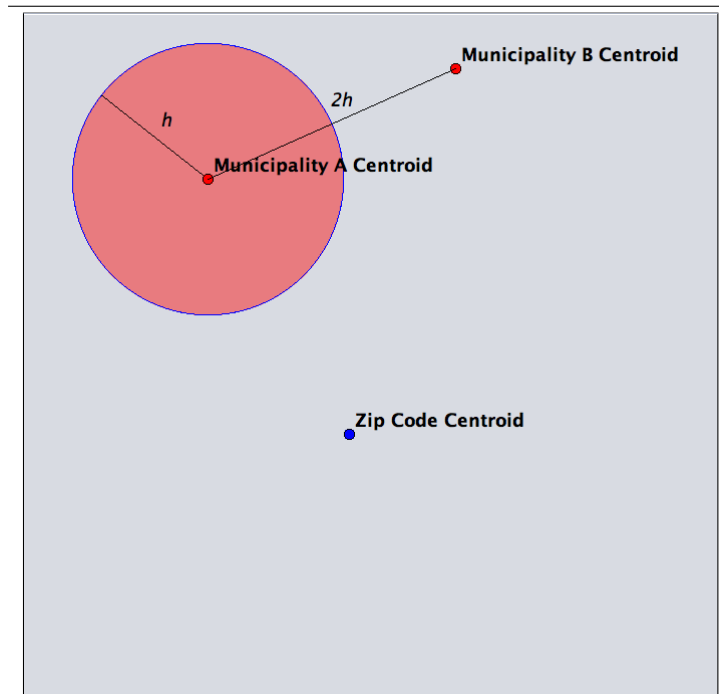
<sup>45</sup>In contrast to the rest of the paper (see footnote 15), the word firm here refers to groups of establishments with the same owner. If a worker has employment relationships with two different establishments within the same firm, only one of them will thus be able to count as a new hire out of unemployment.

<sup>46</sup>In cases where a worker has multiple employment relationships in a given year and where one or more of them could have been preceded by at least three weeks of unemployment, employment relationships are ordered in terms of length and then sequentially flagged as reflecting a new hire out of unemployment starting with the longest spell. This process then continues until the longest remaining employment spell could not have been preceded by three weeks of unemployment if all the employment spells already flagged as new hires out of unemployment were preceded by three weeks of unemployment.



the shortest travel time to municipality A and place the zip code centroid, the municipality A centroid and the municipality B centroid on a two-dimensional plane by treating their pair-wise travel times as the Euclidean distances between them. Next denote the travel time between municipality A and its closest neighbor, municipality B, by  $2h$  and place a disc with radius equal to half this distance,  $h$ , around the municipality A centroid. Note that since larger municipalities will generally have a closest neighbor that is further away, larger municipalities will here have a larger disc around their centroid. Figure 9 illustrates the situation.

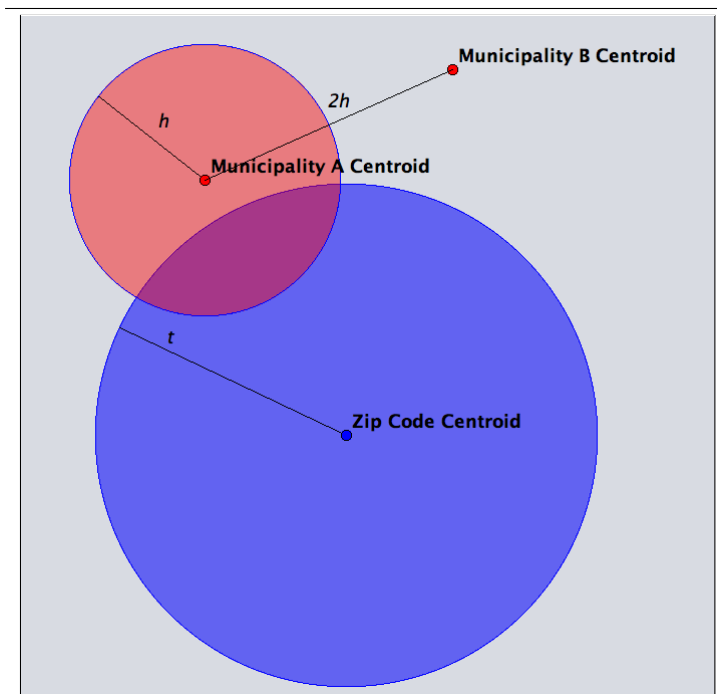
Figure 9: Measuring local labor market size I



The figure illustrates the first step of computing how many unemployed workers and vacancies from municipality A are within  $t$  minutes of the zip code. The centroids of municipality A, its closest neighboring municipality, municipality B, and the zip centroid in question has been placed on a two-dimensional plane by treating their pair-wise travel times as the Euclidean distance between them. A red disc with radius equal to half the distance between municipality A and B has then been placed around the municipality A centroid.

Next, I proceed as if all the unemployed workers and vacancies in municipality A are uniformly distributed on the disc around the centroid. To figure out what share of these are within  $t$  minutes of the zip code centroid we can then simply place a disc with radius  $t$  around the zip code centroid and compute what share of the disc around municipality A intersects with the disc around the zip code centroid. Figure 9 illustrates the situation.

Figure 10: Measuring local labor market size II



The figure illustrates the second step of computing how many unemployed workers and vacancies from municipality A are within  $t$  minutes of the zip code. The centroids of municipality A, its closest neighboring municipality, municipality B, and the zip centroid in question has been placed on a two-dimensional plane by treating their pair-wise travel times as the Euclidean distance between them. A red disc with radius equal to half the distance between municipality A and B has then been placed around the municipality A centroid. Finally a blue disc with radius  $t$  has been placed around the zip code centroid. Assuming the unemployed workers and vacancies in municipality A are uniformly distributed on the red disc, the size of the intersection between the blue and red disk is then used to determine how many of them are within  $t$  minutes of the zip code centroid.

Knowing what share of the vacancies and unemployed workers from municipality A is within  $t$  minutes of the zip code centroid it is easy to figure out the total number of vacancies and unemployed workers from municipality A that are within  $t$  minutes of the zip code centroid. Repeating this for all other municipalities and summing the yields the total number of unemployed workers and vacancies that are within  $t$  minutes.