Mismatch Unemployment and the Geography of Job Search: Online Appendix

By Ioana Marinescu and Roland Rathelot*

Could we significantly reduce U.S. unemployment by helping job seekers move closer to jobs? Using data from the leading employment board CareerBuilder.com, we show that, indeed, workers dislike applying to distant jobs: job seekers are 35 percent less likely to apply to a job 10 miles away from their ZIP code of residence. However, because job seekers are close enough to vacancies on average, this distaste for distance is fairly inconsequential: our search and matching model predicts that relocating job seekers to minimize unemployment would decrease unemployment by only 5.3 percent. Geographic mismatch is thus a minor driver of aggregate unemployment.
JEL: E24, J21, J61, J62, J64.
Keywords: local labor markets, job search, misallocation, mismatch, applications, vacancies, unemployment

Job seekers' optimal application strategy

Here, we derive job seekers' optimal strategies. Let $\mathbf{v} = \{v_1, \dots v_{\bar{a}}\}$ be the $\bar{a}$-tuple of vacancies worker u applies to. We use the convention that utilities are ranked as: $w_{uv_1} \geq w_{uv_2} \geq \dots w_{uv_{\bar{a}}}$. The expected utility associated with strategy $\mathbf{v}$ is:

$$\text{(A1)} \qquad \mathcal{U}(\mathbf{v}) = \pi_{j(v_1)} w_{uv_1} + \sum_{k=2}^{\bar{a}} \left[\prod_{\ell=1}^{k-1} (1 - \pi_{j(v_\ell)}) \right] \pi_{j(v_k)} w_{uv_k}$$

With probability $\pi_{j(v_1)}$, the job seeker u gets an offer from the highest utility vacancy v_1, which is located in j. Whatever other offers he might get, he takes v_1 and his utility is w_{uv_1}. He only takes an offer from vacancy v_k if he does not get

* Marinescu: ioma@upenn.edu; Rathelot: r.rathelot@warwick.ac.uk. We would like to thank Pierre Cahuc, Raj Chetty, Bruno Crépon, Xavier D'Haultfoeuille, François Fontaine, Florence Goffette-Nagot, Larry Katz, Philip Kircher, Francis Kramarz, Thomas Le Barbanchon, Nathan Hendren, Etienne Lehmann, Manasa Patnam, Barbara Petrongolo, Julien Prat, Craig Riddell, Thijs van Rens, Etienne Wasmer and participants to seminars at CREST, Cachan/Paris-Sud, GATE (Lyon), Mannheim, Uppsala, Louvain-la-Neuve, Bonn, EIEF, FGV (Rio), PUC (Rio), Sciences Po, Stockholm, Vienna, Warwick, Aarhus, Copenhagen, Montreal, Pompeu Fabra, UIC, Harvard, Harris School (Chicago), Booth (Chicago), Queen Mary University, Helsinki, George Washington University, Norges Bank, as well as the SaM Meeting at Edinburgh, the ESEM at Toulouse, the workshop Topics in Labor Economics, the NBER LS Spring Meeting at San Francisco 2015, SOLE/EALE world meetings Montreal 2015 for useful comments. We are grateful to CareerBuilder.com for letting us use their data. Any opinions expressed here are those of the authors and not of any institution.

any offer from higher utility vacancies $v_{k'}, k' < k$, which happens with probability $\prod_{\ell=1}^{k-1}(1 - \pi_{j(v_\ell)})$.

Determining which strategy maximizes expected utility in equation A1 is complex: an algorithm such as the one described in **?** should be used. In the general case, it is *not* an optimal strategy to apply to the $\bar{a}$ highest expected utility jobs. Instead, workers should first apply to the highest expected utility job, and then gamble upwards by applying to jobs that have lower probability of yielding an offer but higher utility. Computing the optimal strategy using the **?** algorithm would make our model computationally intractable. We must therefore find some reasonable simplifying assumption to restore tractability.

One way to simplify the problem is to assume that the probability of a worker getting more than one offer is zero. **?** assume that the probability of getting an offer from any given job is so low that the probability to receive two offers or more is negligible. In this case, the expected utility simplifies to $\mathcal{U}(\mathbf{v}) = \sum_k \pi_{j(v_k)} w_{uv_k}$, implying that the optimal strategy is to apply to the vacancies with the highest expected utility.

Another way of simplifying the problem is to assume that the probability π_v of getting an offer and the utility w_{uv} associated with a vacancy v are not negatively correlated. In this particular case, applying to the $\bar{a}$ vacancies with the highest expected utility is optimal, and the model becomes computationally tractable. The intuition is this: if the probability of getting the job and the reward are not negatively correlated, there is no trade-off between risk and reward (utility), and there is therefore no opportunity for gambling upwards. Therefore, if there is no negative correlation between the probability of getting an offer from a job in a location j and the utility derived from a job in location j, it is optimal to apply to the highest expected utility vacancies.

How likely is it that there is no negative correlation between the probability of getting an offer from a job in a location j and the utility derived from a job in location j? Utility is the product of two terms: $f(d)$ is strictly decreasing with geographic distance and ε is an idiosyncratic shock. By assumption, ε is a random draw across vacancies, and thus will not generate any correlation between the probability of getting an offer π and the utility w for a given vacancy.

Then, only a positive correlation between the probability of getting an offer π and the distance d may generate a negative (remember that $f(d)$ is strictly decreasing in d) correlation between the probability of getting an offer and utility. Unfortunately, it is hard to directly measure the correlation between π and the distance d because we don't observe the probability of getting an offer but instead infer it on the basis of applicants' behavior. Therefore, the inferred probabilities of getting an offer π_j in different locations j depend precisely on the assumption about the strategy pursued by job seekers. To make the case that the correlation is unlikely to be negative, we use two arguments. First, we show that, based on the structure of the problem and the data, the correlation between a job's utility and the probability of getting an offer is unlikely to be strongly negative.

Second, we use the fact that, in the hires-maximizing allocation of job seekers, the correlation between the probability of getting an offer π and the distance d is zero and therefore non-negative.

Using the first line of argument, we can say that, in general, if job seekers are geographically dispersed as is the case in our data, π and d cannot be highly correlated either positively or negatively. To see this, suppose that there are only two places A and B, and two job seekers X and Y who live respectively in A and B. Jobs in place A have a higher probability π of generating an offer than jobs in place B. Therefore, for job seekers like X, there is a negative correlation between distance and the probability of getting an offer. For job seekers like Y, there is a positive correlation between distance and the probability of getting an offer. So, depending on the job seekers' location, the correlation between distance and the probability of getting an offer from a job could be positive or negative, implying that overall the correlation cannot be strongly positive or negative.

The question then becomes: how frequent are job seekers like Y and how often do opportunities for gambling upwards arise? In the simple example above, the opportunity for gambling upwards only arises if, for job seeker Y, jobs in A have a higher expected utility than jobs in B. In this case, job seeker Y would not only apply to jobs with the highest expected utility in A, but would want to gamble upwards by applying to jobs in their own location B that have a higher utility but a lower probability of yielding an offer. For jobs in A to have a higher expected utility than jobs in B for Y, it must be that the distance from B to A is not too large and/or that the probability of getting an offer from a job in A is large enough. More generally, this suggests that applying to the highest expected utility jobs is not optimal for job seekers in places where the probability of getting an offer increases more steeply with distance than the disutility of distance.

The conclusion of this first line of argument based on the structure of the problem and the data is this: as long as there are few job seekers for whom labor market conditions (as measured by the probability of generating an offer π) improve drastically within 60 miles or so of their place of residence (remember than 90 percent of application are sent within 60 miles), the assumption that the probability of getting an offer π and utility are not negatively correlated will be generally correct.

The second line of argument relies on the hires-maximizing allocation of job seekers. In this allocation, π_j is equal across all locations j (see equation B2): therefore, there is no correlation between the probability of getting an offer π and distance d, and so applying to the highest expected utility vacancies is indeed optimal. Since it turns out that the actual allocation of job seekers is fairly close to the hires-maximizing allocation of job seekers (there is little mismatch), the π_j tend to be very similar across locations, and there is therefore not much correlation between the probability of getting an offer π and distance d. In conclusion, the assumption that there is no negative correlation between the probability of getting an offer π and distaste for distance $f(d)$ seems reasonable given the structure of the problem and the fact that the allocation of job seekers is close to the hires-maximizing allocation.

Number of matches when job seekers have no distaste for distance

Starting from equation (3), we examine the case in which job seekers have no distaste for distance, i.e. $g(d_{ij}) = 1, \forall i, j$. We derive the probability for a job seeker in i to apply to a vacancy in j p_{ij} as:

$$p_{ij} = \bar{a}\frac{\pi_j^\alpha}{\sum_\ell \pi_\ell^\alpha V_\ell}, \forall i, j \tag{B1}$$

In this case, p_{ij} does not depend on i. Let $\bar{U}$ and $\bar{V}$ be the total number of job seekers and vacancies in the economy. We now derive the probability of getting an offer π. We have, for all j:

$$\begin{aligned}\pi_j &= q\mathcal{R}\left(q\bar{a}\pi_j^\alpha \sum_k \frac{U_k}{\sum_\ell \pi_\ell^\alpha V_\ell}\right)\\ &= q\mathcal{R}\left(\bar{U}\frac{q\bar{a}\pi_j^\alpha}{\sum_\ell \pi_\ell^\alpha V_\ell}\right)\end{aligned}$$

The only term that depends on j on the right-hand side is π_j itself. Therefore, solving for π_j is the same for any ZIP code j. Hence π is equal across ZIP codes in the case of no distaste for distance. Since π is equal across ZIP codes, we can rewrite π as a function of parameters, i.e.:

$$\pi = q\mathcal{R}\left(q\bar{a}\frac{\bar{U}}{\bar{V}}\right) \tag{B2}$$

If $g(d_{ij}) = 1$, the total number of matches is:

$$M = \sum_k U_k \left[1 - \exp\left(-\bar{a}\frac{\sum_\ell \pi_\ell^{1+\alpha} V_\ell}{\sum_\ell \pi_\ell^\alpha V_\ell}\right)\right] \tag{B3}$$

Since π is equal across ZIP codes, the total number of matches when there is no distaste for distance is:

$$M = U\left[1 - \exp\left(-\bar{a}\pi\right)\right]$$

Replacing π by its expression in equation (B2),

$$M = \bar{U}\left[1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\frac{\bar{U}}{\bar{V}}\right)\right)\right] \tag{B4}$$

Thus, the number of matches obtained with no distaste for distance depends on the aggregate number of job seekers $\bar{U}$ and the inverse of aggregate labor market tightness $(\bar{U}/\bar{V})$. Since there is no distaste for distance, only the aggregates matter: the location of jobs and job seekers is irrelevant. The total number of matches also depends on $q\bar{a}$, i.e. the product between the probability of a valid application and the average number of applications sent by a job seeker, which is equal to the average number of valid applications per job seeker. This makes sense since, intuitively, a larger number of valid applications leads to more matches.

Mismatch unemployment by education

Our main results assume that job seekers are homogeneous: here we estimate mismatch while allowing for worker heterogeneity by education. Specifically, we divide job seekers in three educational groups: high school graduates, associate degrees (AA), and bachelor degrees (BA) and more.[1] We also compute the number of vacancies for each education category based on the SOC code of each vacancy and O*NET's determination of the level of education needed in each SOC code.

We compute mismatch by education assuming that job seekers only apply to jobs in their own educational category, so that each education level is a completely separate market. In a first version, we keep all parameters as in the baseline case (i.e. Table 2), except for the geographic distribution of job seekers and vacancies. Mismatch decreases with education (Figure E2). Yet, even for high school graduates, mismatch is only 6.9 percent. In a second version, we adjust all parameters for each education category, and we find that mismatch for high school graduates and AA is only about 4 percent, while mismatch for BA and above is only 1.8 percent (Figure E2).[2]

Overall, since mismatch remains low even for less educated workers, these results reinforce our main conclusion that geographic mismatch is a minor driver of U.S. aggregate unemployment.

[1]In our data, we cannot separate high school dropouts from individuals with missing information on education. While mismatch is likely to be higher for high-school dropouts than for high-school graduates, we cannot estimate a mismatch index for this category.

[2]See Tables E2 and E3 in appendix for the parameters used and for the estimated distaste for distance parameters by education.

A simpler mismatch index, and applications

In this section of the appendix, we investigate how mismatch varies with the Pareto parameter for the match-specific utility component α, and we show that a simpler mismatch index can be derived when $\alpha = 0$. Finally, we show how this simpler mismatch index can be used to compute mismatch under less restrictive hypotheses while preserving computational feasibility.

D1. Calculating a simple mismatch index

Specifically, we vary α between 0 and 2 in increments of 0.2 (remember that our baseline estimate is $\alpha = 0.4629$). Mismatch is maximum at 6 percent when $\alpha = 0$ and decreases for larger values of α (appendix Figure E3). This makes sense because α can be interpreted as the weight put by applicants on the probability of getting an offer from a given vacancy relative to the distance to that vacancy (see equation 3). A smaller α increases mismatch because it hinders job seekers from directing applications to vacancies with higher probability of yielding an offer. Since we estimate a value of α that is close to 0, our geographic mismatch is close to the maximum that it could be as a function of α.

When $\alpha = 0$, job seekers only care about distance and do not take into account the probability of getting an offer when they apply, i.e. they are not strategic. In this case, the mismatch index simplifies considerably because we do not need to ensure that the probability of getting an offer π is consistent with the behavior of job seekers as was the case in equation (5). The mismatch takes a closed form that depends only on where job seekers and vacancies are located and job seekers' distaste for distance:

$$\mathcal{M}_{ns} = 1 - \sum_k \frac{U_k}{M^* \bar{U}} \left[1 - \exp\left(-q\bar{a} \frac{\sum_\ell g(d_{k\ell}) V_\ell \mathcal{R}(q\bar{a}\nu_\ell)}{\sum_\ell g(d_{k\ell}) V_\ell} \right) \right] \tag{D1}$$

where $\mathcal{R}(x) = [1 - \exp(-x)]/x$, M^* is defined in equation (7) and ν_j is a generalized inverse tightness[3] in the no-strategy case defined as:

$$\nu_j = \sum_k \frac{g(d_{kj}) U_k}{\sum_\ell g(d_{k\ell}) V_\ell} \tag{D2}$$

Mismatch with non-strategic job seekers is very similar but slightly higher than our baseline estimates (compare appendix Figure E4 and Figure 6 interconnected). This is not surprising since job seekers do not behave optimally:

[3] If we are interested in measuring the number of job seekers who compete for a job in a ZIP code j, we don't want to use the simple inverse tightness U_j/V_j because job seekers apply to jobs beyond their own ZIP code. Since labor markets are interconnected, the generalized inverse tightness at a place j will depend on the number of job seekers and job vacancies around j. To illustrate how the generalized inverse tightness ν_j varies with j, we plot it for each ZIP code j in the U.S. (appendix Figure E5).

they apply to vacancies only as a function of distance, and do not take into account the probability of getting an offer. Overall, we conclude that, in the case of the U.S. in 2012, this mismatch index with non-strategic job seekers is a fair approximation of our more comprehensive approach.

Because it is much simpler to compute, this non-strategic mismatch index could be straightforwardly used to calculate mismatch with other datasets that contain the geographic distribution of job seekers and vacancies, U_i, V_j. Apart from the distribution of job seekers and vacancies, only two other ingredients are needed:

- The distaste for distance g, which we provide in Table 1. Alternatively, users can specify any other distaste for distance.
- $q\bar{a}$, the scale parameter, which should be calibrated using a target job finding rate.

Mismatch is maximum when $\alpha = 0$, but it is still only 6 percent. Furthermore, the assumption that $\alpha = 0$ yields a simpler mismatch index that can be used in other applications.

D2. Mismatch when employers have a distaste for hiring distant workers

In our main analysis, we assume that employers do not differentiate between workers on the basis of distance, so that the job finding rate per application q does not depend on the distance between the job seeker and the job. Here we relax this assumption, and we use the simple mismatch index expression to do so.

The data does not allow us to separately identify the distaste for distance for employers and job seekers. We therefore let the distaste for distance for employers take different values: zero (our baseline case), half of the distaste for distance that workers exhibit, and the same distaste for distance as workers. Employers' distaste for distance is unlikely to be as strong as workers' distaste for distance because workers would typically bear most of the moving and commuting costs. Since mismatch increases with employers' distaste for distance, assuming that employers have the same distaste for distance as workers is likely to yield an upper bound for mismatch.

Specifically, in the simple mismatch index, we allow $q\bar{a}$, the job finding rate multiplied by the average number of applications, to differ with d_{ij}, the distance between worker i and job j.

The job-finding rate for a worker located in i is equal to:

$$r_i = 1 - \exp\left(\frac{\sum_\ell \bar{a}q(d_{i\ell})g(d_{i\ell})V_\ell \mathcal{R}(\tilde{\nu}_\ell)}{\sum_\ell g(d_{i\ell})V_\ell}\right)$$

where $\tilde{\nu}_\ell = \sum_k \frac{\bar{a}q(d_{k\ell})g(d_{k\ell})U_k}{\sum_j g(d_{kj})V_j}$ is a modified version of our generalized inverse tightness, which accounts for the fact that employers value less the applications coming from further away.

We specify $\bar{a}q(d) = \bar{a}qg(d)^{\zeta}$, with $g(d)$ being the distaste for distance of workers, and ζ a parameter indicating how much employers dislike applications from far away. In what follows, we experiment with different values of ζ and calibrate the $\bar{a}q$ to get the right job finding rate overall. In detail, we wish to find the $q\bar{a}$ that minimizes:

$$\left[\frac{\sum_i U_i r_i(q\bar{a})}{\sum_i U_i} - \bar{r}\right]^2$$

Once $q\bar{a}$ is known, we can compute the total number of matches:

$$M = \sum_i U_i r_i(q\bar{a})$$

as well as the maximum number of matches for the same job seekers:

$$M^* = \bar{U}\left[1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\frac{\bar{U}}{\bar{V}}\right)\right)\right]$$

where $\bar{U}$ is the total number of job seekers.

At the zip code level, the mismatch index is equal to 5.7 percent, 8.5 percent and 8.5 percent respectively when employers have no distaste for distance, have half workers' distaste for distance, and have the same distaste for distance as workers. As is intuitive, mismatch increases when employers also exhibit a distaste for distance, and 8.5 percent is likely to be an upper bound for mismatch when employers dislike applications from far away.

D3. Geographic mismatch with a different q in each occupation

In this subsection, we relax the assumption that the scale parameter q is constant across occupations. We extend the geographic mismatch index from section III.A by allowing q to depend on the previous occupation of job seekers, at the 2-digit SOC level.

We use the Current Population (CPS) basic extracts to compute empirical job finding rates by 2-digit SOC code as follows:

1) Create a panel data using basic CPS April through July 2012.

2) Use a crosswalk to convert the CPS occupation codes into SOC 2010 codes.

3) Calculate the job finding rate by 2-digit SOC of origin and by month, using weights.

4) For each 2-digit SOC, take the average across months.

The objective is to allow $q\bar{a}$ in the simple mismatch index to be different for each 2-digit SOC occupation and to reflect the job finding rates of each occupation.

The job finding rates in each occupation are used as targets to estimate $q\bar{a}$ in this task. We denote these targets as $\bar{r}^o$, for occupation o, and q^o the q specific to the previous occupation o of job seekers.

The job-finding rate for a worker in occupation o and located in i is equal to:

$$r_i(q^o\bar{a}) = 1 - \exp\left(-q^o\bar{a}\frac{\sum_\ell g(d_{i\ell})V_\ell \mathcal{R}(q^o\bar{a}\nu_\ell)}{\sum_\ell g(d_{i\ell})V_\ell}\right)$$

where $\nu_\ell = \sum_k \frac{g(d_{k\ell})U_k}{\sum_j g(d_{kj})V_j}$ is our generalized inverse tightness.

Calling U_i^o the number of workers of occupation o in location i, we wish to find the $q^o\bar{a}$ that minimizes:

$$\left[\frac{\sum_i U_i^o r_i(q^o\bar{a})}{\sum_i U_i^o} - \bar{r}^o\right]^2$$

Once $q^o\bar{a}$ is known, we can compute the total number of matches:

$$M(o) = \sum_i U_i^o r_i(q^o\bar{a})$$

as well as the maximum number of matches for the same job seekers:

$$M^*(o) = \bar{U}^o\left[1 - \exp\left(-q^o\bar{a}\mathcal{R}\left(q^o\bar{a}\frac{\bar{U}}{\bar{V}}\right)\right)\right]$$

where $\bar{U}$ is the total number of job seekers (from all occupations).

The overall mismatch is then:

$$1 - \frac{\sum_o M(o)}{\sum_o M^*(o)}$$

At the county level, mismatch is equal to 5.5 percent when we allow $q\bar{a}$ to vary across occupations,[4] to be compared to 6.6 percent when we ignore the variability of $q\bar{a}$. Therefore, allowing q to vary across occupations does not substantively impact the measure of mismatch in the economy.

Because we are using the simpler mismatch index, we can even calculate mismatch at the zip code by occupation level. Mismatch is equal to 5.3 percent when we allow $q\bar{a}$ to vary across occupations to be compared to 6.4 percent when we ignore the variability of $q\bar{a}$. Therefore, using this simpler mismatch index also allows us to show that the calculation at the county level happens to given levels of mismatch that are similar to what is measured at the zip code level.

[4]For a few occupations, it is not possible to find a $q\bar{a}$ high enough that the job finding rate $r_i(q\bar{a})$ reaches the empirical job finding rate. Since, in practice, mismatch decreases with a higher level of $q\bar{a}$, it is likely that this issue leads us to overestimate mismatch.

Additional figures and tables

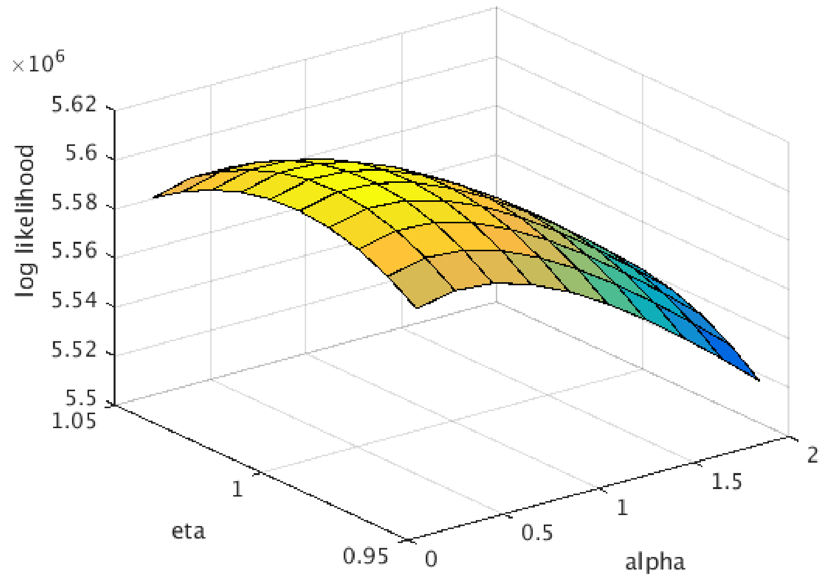

Figure E1. Log likelihood as a function of η, the scaling parameter for the distaste for distance, and α the Pareto parameter for the match-specific utility shock

Source: CareerBuilder database.

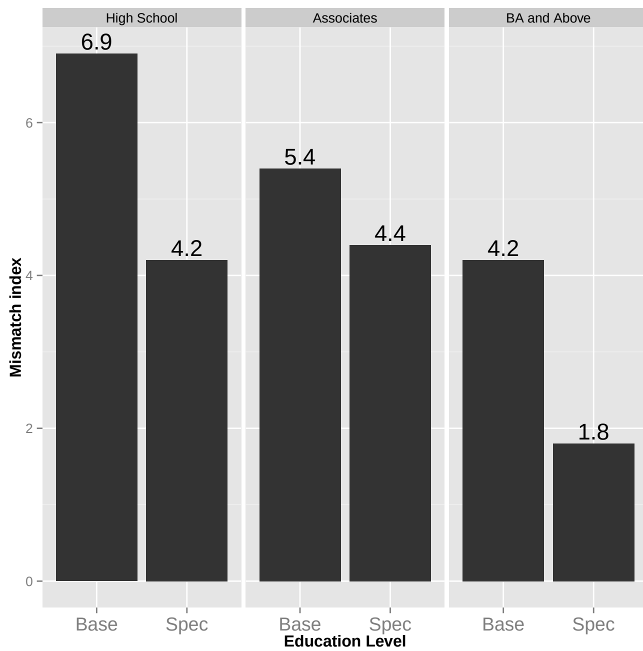

Figure E2. Mismatch unemployment by education: baseline parameters ("Base") and each education category's own specific parameters ("Spec")

Source: CareerBuilder database.

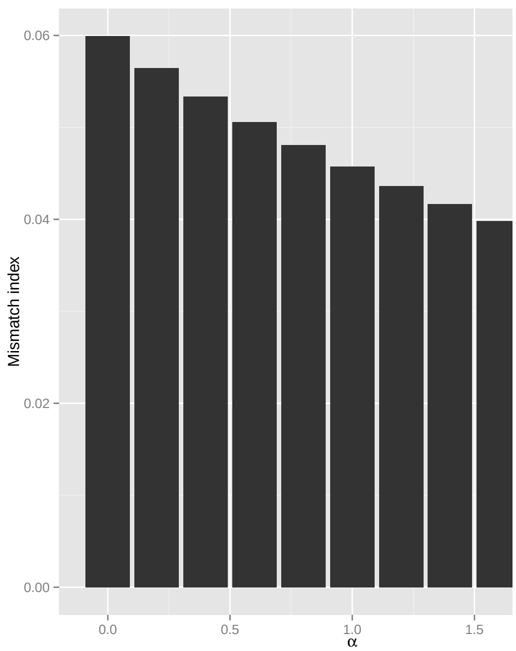

Figure E3. Robustness to various values of the Pareto parameter for the match-specific utility component α

Source: CareerBuilder database.

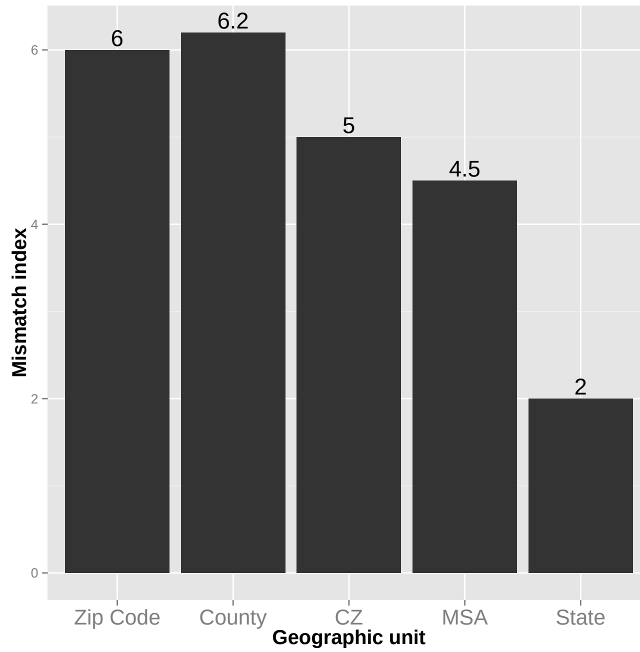

Figure E4. Mismatch unemployment with interconnected markets and non-strategic job seekers

Source: CareerBuilder database.

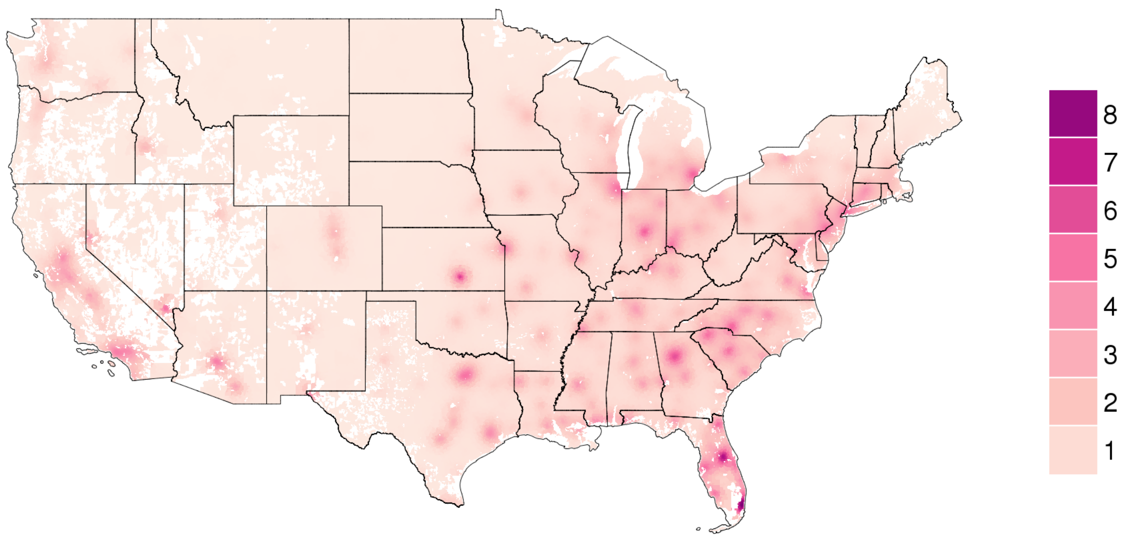

Figure E5. Generalized inverse tightness: number of unemployed workers per job, taking into account the geography of job search

Source: CareerBuilder database.

Table E1— Estimation of the $CZ \times SOC$ model

	(1)	(2)	(3)
Geographic distance			
< 50 miles	-0.0405	-0.0711	-0.0671
	0.0027	0.0021	0.0025
< 75 miles	0.0190	0.0586	0.0582
	0.0075	0.0068	0.0080
< 100 miles	-0.0573	-0.0564	-0.0603
	0.0115	0.0098	0.0117
< 200 miles	0.0598	0.0459	0.0463
	0.0072	0.0061	0.0073
< 500 miles	0.0155	0.0193	0.0194
	0.0011	0.0014	0.0016
< 1,000 miles	0.0031	0.0032	0.0030
	0.0002	0.0003	0.0003
< 2,000 miles	0.0001	-0.0003	-0.0004
	0.0001	0.0002	0.0002
> 2,000 miles	0.0004	0.0022	0.0021
	0.0001	0.0003	0.0003
SOC2			
Different SOC2	-1.2922	-1.0231	-0.7224
	0.0696	0.0539	0.0437
Distance SOC2	-0.2271	-0.3734	-0.4532
	0.0261	0.0223	0.0113
Difference Factor 1	0.2957	0.6082	0.0061
	0.0112	0.0130	0.0119
Difference Factor 2	0.2598	0.3551	0.1916
	0.0121	0.0116	0.0100
N	83,533,150	80,833,282	67,653,650
Fixed-effects	No	User CZSOC	Job CZSOC

Notes: Poisson model (column 1) or Conditional Fixed-Effect Poisson model (columns 2 and 3). Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
The 10 nodes for the spline that parametrizes workers' willingness to apply as a function of distance are at 10, 20, 30, 50, 75, 100, 200, 500, 1000 and 2000 miles. The piecewise-linear spline function is defined by its slopes. With 10 nodes $\{\bar{d}_i\}_{i=1...10}$, the spline is parameterized by 11 parameters $\{\gamma_i\}_{i=1...(11)}$. It is defined so that the derivative of the spline with respect to distance is $s'(d) = \gamma_1$ when distance is below the first node, i.e. when $d < \bar{d}_1$; $s'(d) = \sum_{i=1}^{j} \gamma_i$ when $d \in (\bar{d}_{j-1}, \bar{d}_j)$ and $j = 2 \ldots 10$; $s'(d) = \sum_{i=1}^{11} \gamma_i$ when $d > \bar{d}_{10}$.
Different SOC2 is a dummy for the SOC2 of the applicant's last job differing from the SOC2 of the vacancy. Distance SOC2 is the distance between the applicant's SOC2 and the vacancy's SOC2. Difference Factor 1 is the difference between the first factor of the applicant's SOC2 and the first factor of the vacancy's SOC2; the same definition holds for Difference Factor 2.

Table E2—Parameters for each education category

Parameter	High School	Associates	BA and above
Number of applications	13.8	14.0	13.6
Tightness	0.20	0.33	0.80
Job Finding Rate	0.17	0.17	0.20

TABLE E3—PROBABILITY OF APPLICATION AS A FUNCTION OF DISTANCE BY EDUCATION: POISSON REGRESSION

	(1) High School	(2) AA	(3) BA and Above
γ_1	-0.0348***	-0.0425***	-0.0498***
	(0.00476)	(0.00222)	(0.00531)
γ_2	-0.0146**	-0.00239	0.00867
	(0.00695)	(0.00338)	(0.00799)
γ_3	-0.000455	0.00112	-0.00425
	(0.00698)	(0.00299)	(0.00736)
γ_4	-0.0461***	-0.0391***	-0.0329***
	(0.00578)	(0.00259)	(0.00705)
γ_5	0.0367***	0.0302***	0.0268***
	(0.00635)	(0.00306)	(0.00756)
γ_6	0.0368***	0.0278***	0.0353***
	(0.00787)	(0.00407)	(0.00827)
γ_7	0.00855	0.0152***	0.00627
	(0.00546)	(0.00287)	(0.00513)
γ_8	0.00974***	0.00500***	0.00631***
	(0.00166)	(0.000822)	(0.00153)
γ_9	0.00401***	0.00431***	0.00346***
	(0.000873)	(0.000334)	(0.000626)
γ_{10}	-0.000335	4.36e-05	-0.000238
	(0.000458)	(0.000230)	(0.000325)
γ_{11}	3.07e-05	0.000253	0.000698***
	(0.000367)	(0.000214)	(0.000237)
Observations	57,997,472	178,134,756	29,997,033
Log-PseudoLikelihood	-122959.6	-1256988.9	-100679.32

Notes: Conditional Fixed-Effect Poisson model with user ZIP code fixed effects. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The 10 nodes for the spline that parametrizes workers' willingness to apply as a function of distance are at 10, 20, 30, 50, 75, 100, 200, 500, 1000 and 2000 miles. The piecewise-linear spline function is defined by its slopes. With 10 nodes $\{\bar{d}_i\}_{i=1...10}$, the spline is parameterized by 11 parameters $\{\gamma_i\}_{i=1...(11)}$. It is defined so that the derivative of the spline with respect to distance is $s'(d) = \gamma_1$ when distance is below the first node, i.e. when $d < \bar{d}_1$; $s'(d) = \sum_{i=1}^{j} \gamma_i$ when $d \in (\bar{d}_{j-1}, \bar{d}_j)$ and $j = 2 \ldots 10$; $s'(d) = \sum_{i=1}^{11} \gamma_i$ when $d > \bar{d}_{10}$.