American Economic Journal: Macroeconomics 2018, 10(3): 42–70
https://doi.org/10.1257/mac.20160312

Mismatch Unemployment and the Geography of Job Search†

By Ioana Marinescu and Roland Rathelot*

Could we significantly reduce US 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 (mi.) 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, J41, J61, J63, J64, R23)

The sharp rise in unemployment during the Great Recession prompted some researchers to consider the role of increasing geographical mismatch between jobs and job seekers. Şahin et al. (2014) found this channel to be of limited importance.[1] Even if geographic mismatch did not contribute much to *increasing* unemployment during the Great Recession, this does not imply that geographic mismatch plays no role in the *level* of unemployment. Determining the level of mismatch unemployment is important because it allows us to predict the effects of policies that aim at bringing job seekers and vacancies closer to each other (Fan 2012; Neumark and Kolko 2010; Kline and Moretti 2014; Busso, Gregory, and Kline 2013; and Neumark and Simpson 2015). In this paper, we exploit a large and rich dataset to determine the level of geographic mismatch in the United States and find that such mismatch is limited.

* Marinescu: University of Pennsylvania School of Social Policy and Practice, 3701 Locust Walk, Philadelphia PA, 19104, and NBER (email: ioma@upenn.edu); Rathelot: Department of Economics, University of Warwick, Coventry, CV4 7AL, United Kingdom, and CEPR (email: 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, and 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.

† Go to https://doi.org/10.1257/mac.20160312 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] Relatedly, evidence does not support the "house lock" hypothesis during the Great Recession. Indeed, homeowners' lower mobility did not contribute to increasing unemployment (Farber 2012, Valletta 2013).

We use zip code level data from CareerBuilder.com on the geography of job search for close to 500,000 job seekers sending more than 5 million applications in 2012. CareerBuilder.com is arguably the largest job board in the United States, and is broadly representative of the US labor market.[2] We find that job seekers are more likely to apply to jobs closer to home: a job seeker is 35 percent less likely to apply to a vacancy that is 10 miles away than to a vacancy that is in the job seekers' zip code of residence. Still, job seekers send 11 percent of their applications to out-of-state vacancies.

To determine the level of geographic mismatch, we use a directed search model where workers strategically choose where to send their applications given that vacancies closer to home yield higher utility. The model predicts where a job seeker applies based on our empirical estimate of distaste for distant jobs, and the expected probability of getting an offer given the locations of employers and all other job seekers. We compute the maximum number of hires that can be obtained by reallocating job seekers across zip codes and leaving the vacancies where they are, which is equivalent to computing the number of hires if job seekers were equally willing to apply anywhere. Finally, geographic mismatch is measured as the difference between the number of hires with the hires-maximizing geographic distribution of job seekers and the number of hires with the existing geographic distribution of job seekers. In fact, 5.3 percent of hires are lost due to job seekers not being close enough to jobs, suggesting that policies aimed at reducing the distance between jobs and job seekers are likely to have a limited impact on *aggregate* unemployment.[3]

Our paper makes two key contributions to the literature. On the theoretical side, we develop a new model of geographic mismatch that fully takes into account the geography of job search, in contrast to prior measures of geographic mismatch that assume job seekers only search in their own location (e.g., Lazear and Spletzer 2012, Şahin et al. 2014, and Herz and van Rens 2015). While Manning and Petrongolo (2017) used a model that takes into account the geography of job search, they did not address mismatch. Our model fully takes into account the geography of job search and thus allows us to pin down the level of geographic mismatch. Finally, we develop a simple mismatch index that is much easier to compute (located in the online Appendix), and yields very similar results to our main estimates.

Our second key contribution is empirical. We use detailed zip code level data on applications in the United States, while Manning and Petrongolo (2017) indirectly infer distaste for distance in the United Kingdom from their model and the location of job seekers and jobs. Our data allows us to take into account applications across geographic units such as metropolitan statistical areas (MSAs), *and* has sufficiently high geographic resolution (zip code level) that we can take into account within MSA frictions as well. Allowing application across geographic units is important because it typically significantly reduces empirical estimates of mismatch. Furthermore, our model and data allow us to combine mismatch by geography and occupation to

[2] Monster.com is the other leading job board and is comparable in size. Which of CareerBuilder or Monster is larger depends on the exact size metric used.

[3] To the extent that such policies create jobs on net, this would change our conclusion, which only pertains to moving jobs or job seekers while keeping their numbers fixed. Furthermore, such policies can have important distributional impacts, which we do not address here.

show that mismatch remains relatively low even when we take into account heterogeneity by two-digit occupations. Arguably, prior literature on geographic mismatch did not attempt to develop models that fully take into account the geography of job search or applications across occupations because of a lack of adequate data to estimate such models. We are thus in the privileged position to have the necessary data to estimate a model of geographic (and occupational) mismatch that is significantly more realistic.

Our paper is related to the literature on mismatch (e.g., Lazear and Spletzer 2012, Şahin et al. 2014, and Herz and van Rens 2015) and the efficiency of the matching function (e.g., Barlevy 2011; Veracierto 2011; Davis, Faberman, and Haltiwanger 2012; and Barnichon and Figura 2015) during and after the Great Recession. Compared to this literature, we focus on precisely measuring one specific type of mismatch: geographic mismatch.

Our paper is related to the literature on geographic mobility in the United States (Molloy, Smith, and Wozniak 2011; Greenwood, Hunt, and McDowell 1986; Bound and Holzer 2000; and Wozniak 2010). Our results complement this literature by investigating the macro effect of such mobility. Our work is also related to the urban economics literature that investigates the distance between the place of residence and the place of employment, and the spatial mismatch hypothesis (Ihlanfeldt and Sjoquist 1998; Hellerstein, Neumark, and McInerney 2008; Rupert and Wasmer 2012; McKenzie 2013; and Guglielminetti et al. 2015). We complement this research with evidence on the job search process.

Finally, the evidence we provide about the geography of job search is relevant to the literature on the impact evaluation of many types of local labor market shocks.[4]

The first section presents the data. In the second section, we present our theoretical framework. In the third section, we provide results about the geography of job search and the level of geographic mismatch. Section IV provides robustness tests and extensions. Section V concludes.

I. The Geography of Job Search

A. *Data*

We use proprietary data provided by CareerBuilder.com, the largest US employment website. We merge three datasets extracted from CareerBuilder's database. The first one is a random sample of registered users whose accounts were active between April and June 2012. For each job seeker, we have the residence location at the zip code level. In order for our results to be comparable with prior

[4] This issue is relevant to measure the impact of immigrants on natives' wages or employment rates (Card 1990; Altonji and Card 1991; Friedberg and Hunt 1995; Borjas, Freeman, and Katz 1996, 1997; Card and DiNardo 2000; Card 2001, 2005; Borjas 2003; amd Ottaviano and Peri 2006), the impact of local shocks on labor demand and supply (Blanchard and Katz 1992, Bound and Holzer 2000, Notowidigdo 2011, and Yagan 2017), the impact of trade and FDI on labor market outcomes (Autor, Dorn, and Hanson 2013, 2015), the equilibrium effects of active labor market policies (Davidson and Woodbury 1993; Blundell et al. 2004; Gautier et al. 2012; Crépon et al. 2013; and Ferracci, Jolivet, and van den Berg 2014), the heterogeneity of the negative duration dependence with local conditions (Kroft, Lange, and Notowidigdo 2013), or spatial mismatch (Patacchini and Zenou 2005; Hellerstein, Neumark, and McInerney 2008; Boustan and Margo 2009; and Åslund, Östh, and Zenou 2010).

literature on job search, we restrict the data to unemployed users. After dropping those who do not reside in the United States, who live in Alaska and Puerto Rico, and those whose location is unknown, we end up with a dataset of 451,783 users.

The second dataset is a sample of vacancies published on the website between April and June 2012, and therefore available to the job seekers to apply to. For each job, we know its location at the zip code level. Removing inconsistent observations, duplicates, and vacancies not located in the United States (or located in Alaska or Puerto Rico), and vacancies without zip code information leaves 696,975 observations. Thirty-seven percent of the vacancy sample is lost due to the zip code availability restriction. We check whether these vacancies without zip code are different in terms of location or occupation compared to the vacancies with a zip code. The correlation between the city counts of vacancies with zip code and without zip code is 0.97. The correlation between the Standard Occupational Classification (SOC) SOC-6 level count of vacancies with and without a zip code is 0.91. We conclude that vacancies without a zip code have the same distribution across cities and occupations as vacancies with a zip code, and thus omitting these vacancies should not bias our results.[5] Finally, the third dataset connects the two previous datasets by showing which jobs each job seeker applied to. An application is defined as a click on the "Apply now" button that can be found on the full job listing web page. On average, job seekers sent around 12.8 applications, and vacancies receive 15.8 applications from job seekers in this sample.

We now address the representativity of the data. Background work (Marinescu and Wolthoff 2015) was done to compare the industry distribution of job vacancies in CareerBuilder.com with the distribution in the Job Openings and Labor Turnover Survey (JOLTS). Compared to the distribution of vacancies across industries in JOLTS, some industries are overrepresented in CareerBuilder data, in particular information technology, finance and insurance, and real estate, rental, and leasing. The most underrepresented industries are state and local government, accommodation and food services, other services, and construction.[6] While the vacancies on CareerBuilder are not perfectly representative of the ones in the US economy as a whole, they form a substantial fraction of the market. Indeed, the number of vacancies on CareerBuilder.com represented 35 percent of the total number of vacancies in the United States in January 2011 as counted in JOLTS.

In terms of occupation (two-digit SOC codes), the distribution of unemployed job seekers' occupations in CareerBuilder data is very similar to the Current Population Survey (CPS) (correlation of 0.71 between the shares of job seekers in each occupation in the two datasets), and the distribution of vacancies' occupations in the CareerBuilder data is essentially identical to the distribution of vacancies in all online jobs (correlation of 0.95 with Help Wanted Online data).

Since the geographic aspect is very important for the purpose of this paper, we verified that the location of vacancies and job seekers in this data is representative

[5] In a robustness test (footnote 28), we include these vacancies in our calculation of mismatch at the MSA and commuting-zone levels, and find that doing so yields almost the same level of mismatch as using the full sample.

[6] This may raise the concern that CareerBuilder data underrepresents some occupations and industries where mismatch may be the highest. In Section IVA below, we compute mismatch for low skilled occupations and low educated workers and show that this mismatch is only slightly higher than the overall mismatch.

of the location of vacancies and job seekers in the United States in general. Across US regions, vacancies in our dataset are distributed very similarly to vacancies in the nationally representative JOLTS in April–June 2012 (96 percent correlation between the shares of vacancies in each region in the two datasets). Across US states, job seekers in this data are also distributed very similarly to the unemployed in the Current Population Survey in April–June 2012, with a correlation of 88 percent.

In our data, job seekers send 11 percent of their applications out of state. Of these, some will commute to the other state, and some will move. Using the American Community Survey (ACS) from 2006 to 2010, we find that 4 percent of employed people commute across state lines for work. Using the 2008 SIPP panel covering years 2008–2013, we find that 5.1 percent of unemployed people who become employed move across states in the six months before and after the event (this number accounts for slight differences in the composition of the SIPP sample in age and education compared to the CareeBuilder sample). These figures added up together are not far from the 11 percent of cross-state applications, which suggests that the vast majority of applications can be considered as "serious."

We also compare the destinations of Americans who move across states in the ACS in 2012 with the destinations of out-of-state applications in our data. We find a very high correlation between the destinations of moves and applications (matrices with the share of moves from each state to each other state), at 0.82. We perform a similar exercise at the county level, comparing the destinations of within-state cross-county applications with within-state cross-county commuting destinations observed in the ACS: we find a high correlation of 0.78.

In conclusion, our data is broadly representative of the US distribution of vacancies and job seekers, and the distribution of applications across geographic units is consistent with the moving and commuting behavior of Americans.

B. *Estimating the Distaste for Distance*

To understand the geography of job search, we must understand how important distance is in job seekers' application behavior. We first use a descriptive approach and show, for each commonly used geographic unit, the share of applications that are sent to jobs within this unit on average across job seekers (Figure 1).[7] The average share of within-state applications is 89 percent. At the other extreme, the average share of applications within zip code is only 4 percent. Overall, this descriptive approach suggests that job seekers are willing to apply away from their zip code but that this willingness declines with distance.

To get a more systematic picture of the impact of geographic distance on job seekers' application behavior, we use a Poisson regression to estimate the probability p_{ij} that a job seeker in zip code i applies to a vacancy in zip code j as a function of distance between i and j. Probability p_{ij} pins down job seekers' distaste for distance, which will be used to calculate the degree of geographic mismatch. The distaste for distance is a shorthand for any preference *or* cost that makes job seekers less likely

[7] All figures have been made with ggplot2 in R (Wickham 2009).

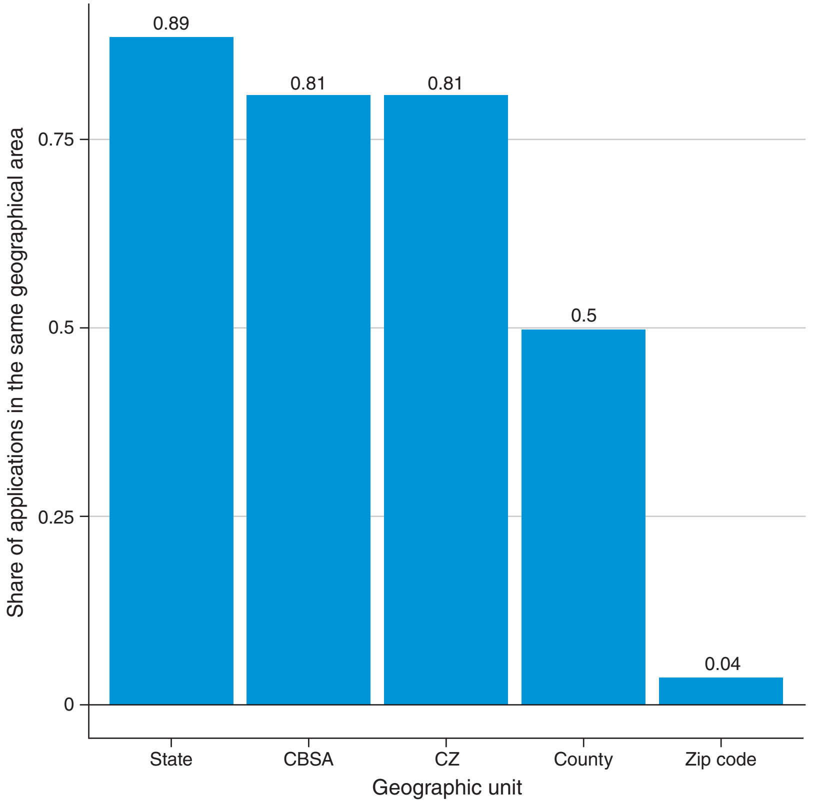

Figure 1. Average Share of Applications Sent within the Same Geographic Area

Source: CareerBuilder.com

to apply to more distant jobs. For example, the distaste for distance may reflect a search cost: it may be harder to gather information about employers who are far away because it is harder to visit workplaces in person.

We model the number of applications from job seekers in zip i to vacancies in zip j as a Poisson[8] with parameter μ_{ij}:

$$\mu_{ij} = U_i V_j \exp\left[\alpha_i + \lambda_j + s(d_{ij})\right], \tag{1}$$

where U_i and V_j are the number of job seekers in i and vacancies in j, α_i and λ_j are fixed effects[9] for job seekers' and vacancies' zip codes, respectively, and $s(\cdot)$ is a spline function whose parameters are estimated. We use a piecewise-linear spline function, defined by its slopes. With n nodes $\{\overline{d}_i\}_{i=1,\ldots,n}$, the spline is parametrized

[8] The data on applications is collapsed by job seeker zip code and vacancy zip code to obtain the total count of applications from i to j.

[9] We estimate conditional fixed effects models to deal with the incidental parameter problem (Hausman, Hall, and Griliches 1984). For the model with two-way fixed effects, we follow the estimation procedure proposed by Guimarães and Portugal (2010) and are only able to perform the estimation on a 10 percent random subsample, given the computational burden.

Table 1—Probability of Application as a Function of Distance: Poisson Regression

	No fixed effect	User fixed effect	Job fixed effect	Two-way fixed effect (10 percent sample)
γ_1	−0.054	−0.047	−0.047	−0.059
	(0.001)	(0.002)	(0.001)	(0.001)
γ_2	0.0005	−0.007	−0.001	0.004
	(0.002)	(0.002)	(0.001)	(0.001)
γ_3	−0.003	−0.010	−0.0001	−0.005
	(0.00158)	(0.002)	(0.002)	(0.001)
γ_4	−0.021	−0.032	−0.034	−0.033
	(0.001)	(0.002)	(0.001)	(0.001)
γ_5	0.016	0.032	0.014	0.030
	(0.001)	(0.002)	(0.002)	(0.001)
γ_6	0.043	0.041	0.044	0.039
	(0.001)	(0.002)	(0.002)	(0.002)
γ_7	0.011	0.014	0.016	0.014
	(0.001)	(0.002)	(0.001)	(0.001)
γ_8	0.005	0.005	0.005	0.005
	(0.000)	(0.000)	(0.000)	(0.000)
γ_9	0.004	0.005	0.005	0.005
	(0.000)	(0.000)	(0.000)	(0.000)
γ_{10}	−0.0002	−0.0002	−0.0003	−0.0004
	(0.000)	(0.000)	(0.000)	(0.000)
γ_{11}	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3.37×10^8	3.37×10^8	3.37×10^8	1.84×10^7
log-Pseudo-likelihood	−7,961,547.5	−6,419,687.7	−6,191,151.4	−487,220.13
Pseudo-R^2	0.7218			

Notes: Poisson model (column 1) or conditional fixed effect Poisson model with user zip code fixed effects (column 2), job zip code fixed effects (column 3) or two-way fixed effects (column 4). Robust standard errors are in parentheses. 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, 1,000 and 2,000 miles. The piecewise-linear spline function is defined by its slopes. With 10 nodes $\{\bar{d}_i\}_{i=1,\ldots,10}$, the spline is parameterized by 11 parameters $\{\gamma_i\}_{i=1,\ldots,(11)}$. 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 $d > \bar{d}_{10}$.

Source: CareerBuilder.com

by $n+1$ parameters $\{\gamma_i\}_{i=1,\ldots,(n+1)}$. 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 $d < \overline{d}_1$; then $s'(d) = \sum_{i=1}^{j} \gamma_i$ when $d \in (\overline{d}_{j-1}, \overline{d}_j)$ and $j = 2, \ldots, n$; $s'(d) = \sum_{i=1}^{n+1} \gamma_i$ when $d > \overline{d}_n$. In other words, for $d < \overline{d}_1$, a 1 mile increase in distance multiplies the probability of application by $\exp(\gamma_1)$. This implies that the probability of application changes by approximately γ_1 *percent* for a 1 mile increase in distance.

We chose 10 nodes $\{\overline{d}_i\}_{i=1,\ldots,10}$ for the spline that parametrizes workers' willingness to apply as a function of distance: at 10, 20, 30, 50, 75, 100, 200, 500, 1,000 and 2,000 miles.[10]

[10] Allowing for a flexible function of distance at smaller distances is important to accurately identify job seekers' distaste for distance. Indeed, we have also experimented with a linear specification in distance and found that it does a worse job than the spline in explaining the data (pseudo $R^2 = 0.53$ versus 0.72 for the spline specification).

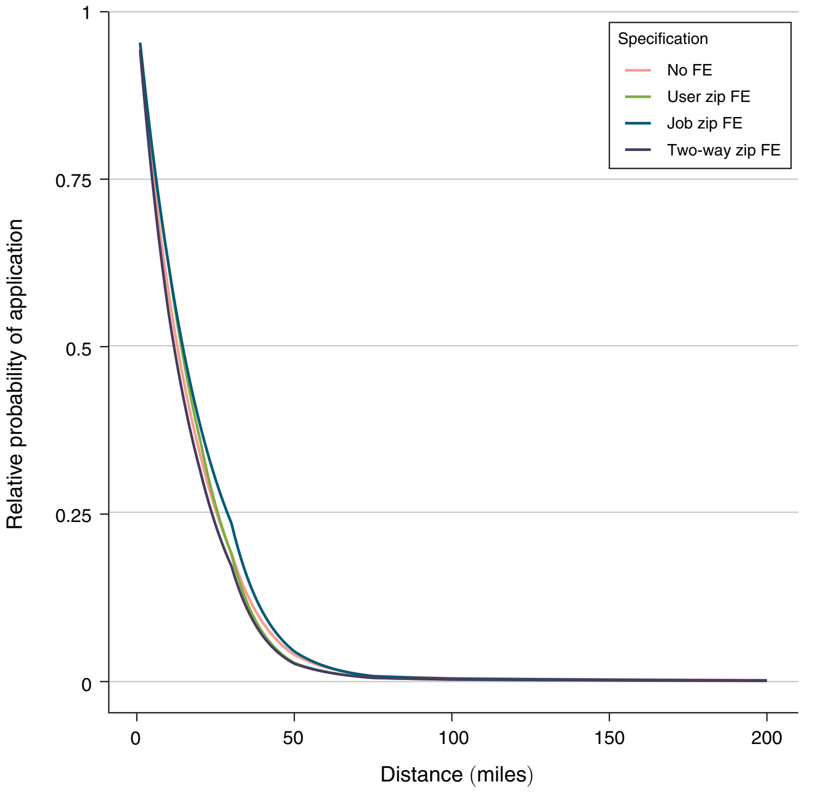

Figure 2. Relative Probability of Application as a Function of Geographic Distance: Predictions from Poisson Model with or without Fixed Effects

Source: CareerBuilder.com

The estimated spline function that captures how far away job seekers apply is displayed graphically in Figure 2, based on the regression coefficients $\{\gamma_i\}_{i=1,...,11}$ in Table 1. Overall, applications clearly decrease with distance. One potential concern is that job seekers in different locations may send different numbers of applications. Similarly, vacancies in some locations may be more attractive to all job seekers, which could bias our estimates of the distaste for distance. Reassuringly, the estimate of the spline is not sensitive to the presence of job seeker zip code and vacancy zip code fixed effects (Figure 2).

Substantively, job seekers are 38 percent less likely to apply to a vacancy 10 miles away than to one in their zip code of residence (estimates without fixed effects in the first column of Table 1). At larger distances, the distaste for distance is much smaller: job seekers are 9 percent less likely to apply to a vacancy 110 miles away from their zip code of residence than to a vacancy 100 miles away. If we take estimates with one-way fixed effects for the job seekers zip codes (the estimates we use to calculate

The linear specification strongly overestimates job seekers' willingness to apply at short distances away from their zip code (under 75 miles) compared to the spline specification.

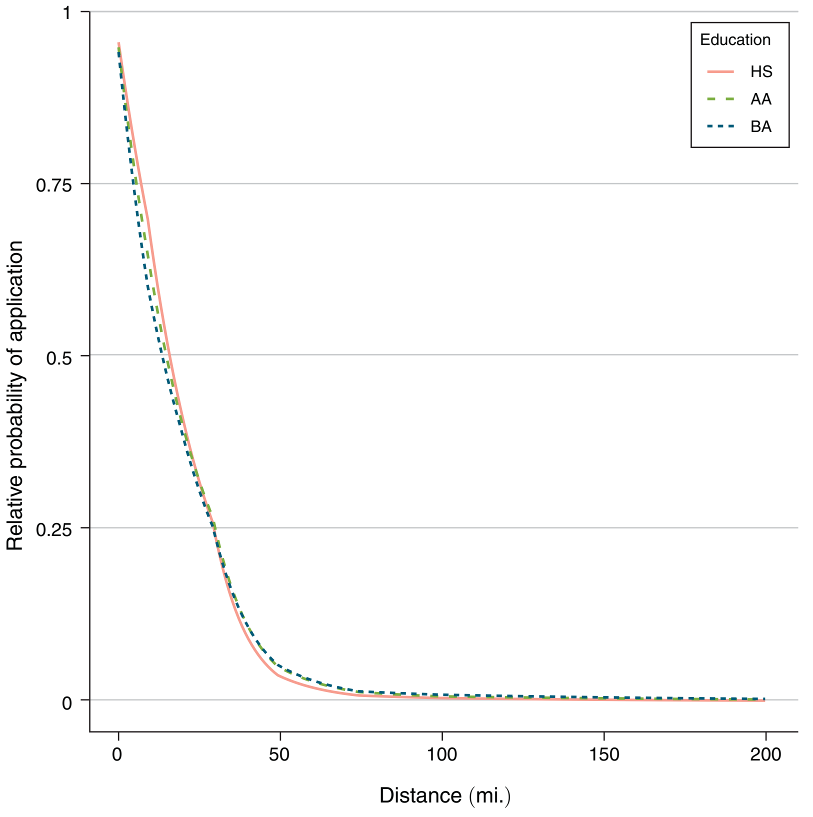

Figure 3. Relative Probability of Application as a Function of Geographic Distance: Predictions from Poisson Model by Education with Job Seeker Zip Code Fixed Effects

Source: CareerBuilder.com

mismatch below), we find very similar effects, with a 35 percent and a 9 percent decline in applications at 10 and 110 miles, respectively.

Is there any systematic difference in the distaste for distance by education or job type? More educated workers are less likely to apply far away from home for short distances (below 30 miles) but more likely to apply far away for long distances (Figure 3). The result for long distances is consistent with the higher mobility of college educated workers across states (Wozniak 2010). In Figure 4, we compare the distaste for distance for the most common 8-digit SOC among job seekers (customer service representatives) and the most common 8-digit SOC among vacancies (registered nurse). Customer service representatives, a relatively low skill occupation, exhibit a higher distaste for distance than the overall sample. However, registered nurses have a higher willingness to apply far away from their zip code of residence than the overall sample.

Overall, we find that job seekers are less likely to apply to vacancies farther away from their zip code of residence, and these results are robust to controls for job seeker and vacancy zip code fixed effects. What is yet to be determined is whether

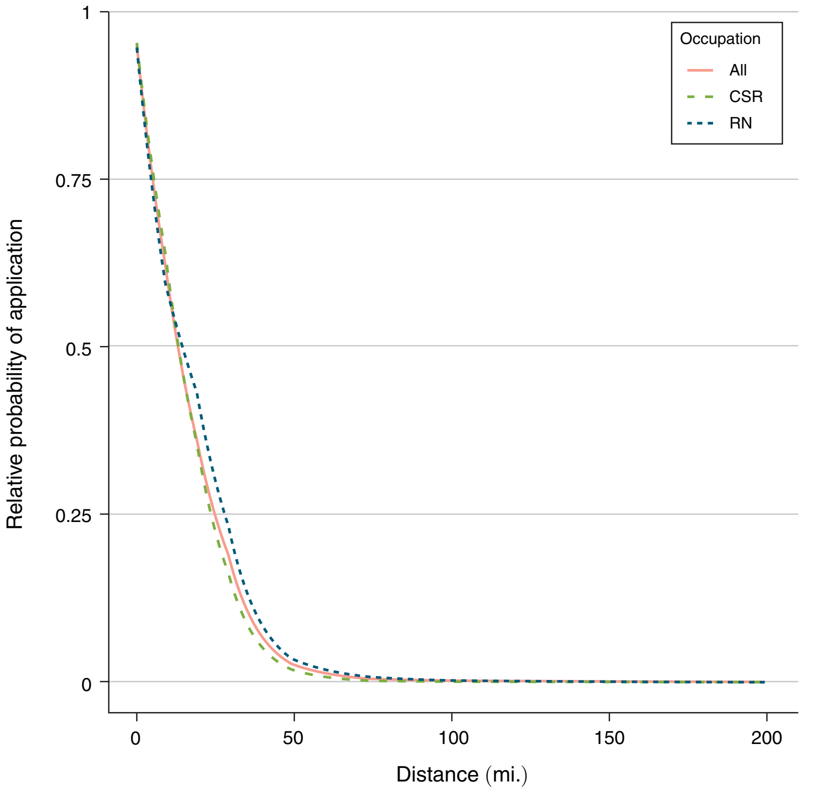

Figure 4. Relative Probability of Application as a Function of Geographic Distance: Predictions from Poisson Model by Education with Job Seeker Zip Code Fixed Effects

Source: CareerBuilder.com

job seekers' preference for jobs close to home is high enough to generate substantial geographic mismatch. This is the topic of the next sections.

II. Mismatch Unemployment with Distinct Labor Markets

Geographic mismatch occurs when there are too many job seekers (relative to jobs) in some places and too few in other places. Therefore, a greater geographic dispersion in labor market tightness (vacancies/unemployment) implies that there is more geographic mismatch. But how can we quantify the impact of a given level of dispersion in tightness on aggregate unemployment? To pin down this impact, we need to make assumptions about how the geographic distribution of job seekers and vacancies affects hires.

Assume that the location of vacancies is exogenous and fixed. Define mismatch as the percent shortfall in hires resulting from the misallocation of job seekers, i.e., 1 − (Total number of hires given observed geographic allocation of job seekers)/(Maximum number of hires across all allocations of job seekers). To calculate the number of hires, we need a matching function, i.e., a mapping from the geographic distributions of U and V to the total number of hires (matches).

The standard approach (Nickell 1982, Jackman and Roper 1987) assumes that job seekers are equally likely to match with any job within their home labor market, and will never match with a job outside their home labor market. A Cobb-Douglas matching function is assumed for each market. In this case, the mismatch index[11] is

$$(2) \qquad \mathcal{M}_{CD} = 1 - \sum_i \left(\frac{V_i}{\sum_i V_i}\right)^{\gamma} \left(\frac{U_i}{\sum_i U_i}\right)^{1-\gamma},$$

where V_i, U_i are the number of vacancies and unemployed workers in geographic area (labor market) i, respectively.

Şahin et al. (2014) shows that the Cobb-Douglas mismatch index $\mathcal{M}_{CD}$ represents the percentage shortfall in hires obtained with the actual allocation of job seekers relative to the hires-maximizing allocation of job seekers. In what follows, we take $\gamma = 0.5$, as in Şahin et al. (2014).

In order to calculate this mismatch index, one must choose a geographic unit for the location of job seekers, such as the MSA. Working with too broad areas is likely to create a downward bias on the index. If there is only one area (e.g., United States), all applications from job seekers residing in this area are obviously sent within the same area. In this case, the index will obviously be equal to zero but will understate the actual geographic mismatch. Conversely, if we use zip codes as the unit of observation, we have the opposite problem. Many applications are directed to vacancies that are not located in the area where the job seeker resides, and we run the risk of overestimating geographic mismatch. As demonstrated by Şahin et al. (2014), choosing a larger area to define the location of job seekers mechanically yields lower mismatch according to the Cobb-Douglas index $\mathcal{M}_{CD}$.

The standard approach assumes that job seekers are as likely to apply to any job within a geographic unit, regardless of how far jobs may be from job seekers' homes. When choosing small units such as zip codes, this assumption seems reasonable. But for larger units such as MSAs, this may no longer be the case, and job seekers may greatly prefer those jobs within the MSA that are closer to home. Therefore, choosing larger search areas will tend to make us underestimate the amount of friction within each geographic unit, and this is a further reason why $\mathcal{M}_{CD}$ is sensitive to the choice of a geographic unit.

Figure 5 shows how mismatch $\mathcal{M}_{CD}$ varies with the size of the geographic area where job seekers are assumed to look for jobs. When job seekers' search area is defined as the state, 1.6 percent of hires are lost due to the misallocation of job seekers. If we define the search area as the MSA or the commuting zone (CZ), mismatch is about 2.5 percent. When search areas are counties, this figure doubles to 4.9 percent; this number is of a similar order of magnitude to the roughly 3 percent mismatch found by Şahin et al. (2014) using county-level data in 2011.[12] At the zip code level, the fraction of hires lost due to misallocation of job seekers is a very large

[11] See Jackman, Layard, and Pissarides (1989); Lazear and Spletzer (2012) for a dissimilarity index, which provides a measure of the proportion of the unemployed who are in the "wrong" market. Using the dissimilarity measure yields qualitative results very similar to Figure 5.

[12] Estimates are not directly comparable due to different data sources and a different time period.

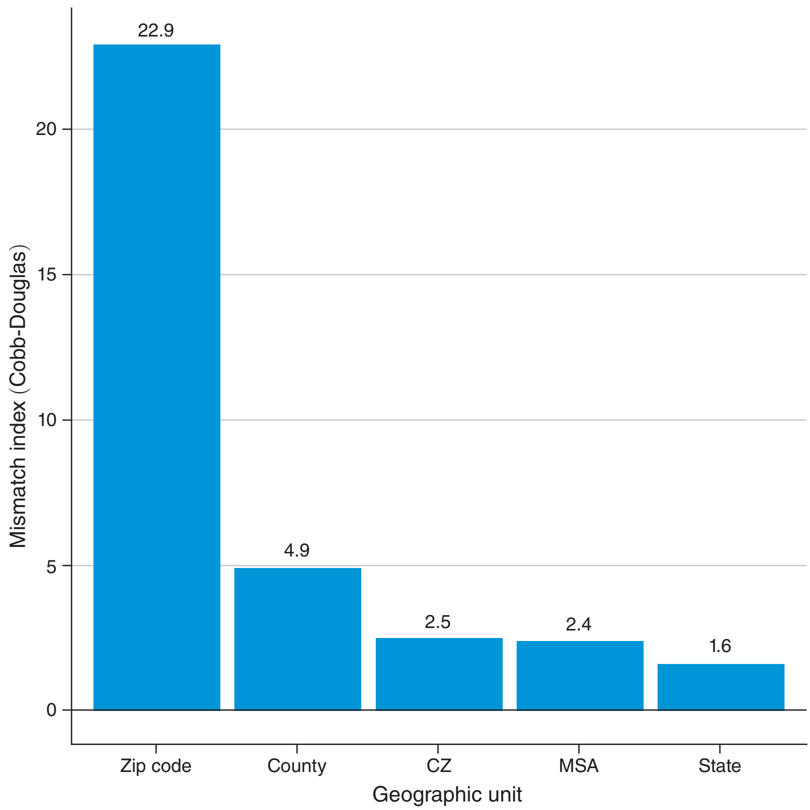

Figure 5. Mismatch Unemployment: Distinct Markets and Cobb-Douglas Matching Function

Source: CareerBuilder.com

22.9 percent. When using the Cobb-Douglas mismatch index $\mathcal{M}_{CD}$, the magnitude of geographic mismatch thus strongly depends on the size of the geographic area where job seekers are assumed to look for jobs, with smaller areas yielding larger mismatch values.

III. Mismatch Unemployment with Interconnected Labor Markets

A. *A Search and Matching Model with Interconnected Markets*

Our approach to mismatch seeks to overcome the limitations of the standard approach, which assumes that job seekers only apply to jobs in their own labor market and are equally likely to apply to any job within their labor market. We modify the standard mismatch index in two ways. First, we allow job seekers to apply to jobs in all locations (zip codes). Second, in order to be able to model applications across locations, we replace the Cobb-Douglas matching function with a standard urn-ball matching function.

Before developing the details of the model, it is worth noting that the model applies to any setting with heterogeneous labor market cells. For example, while our baseline empirical application focuses on heterogeneity by geography, we also apply the model to a case where jobs and job seekers differ by geography *and* occupation.

Our objective is to obtain an expression for the total number of matches as a function of the number of job seekers and vacancies in each location, and structural parameters. We use a directed search model where workers choose where to send their applications based on the *location* of the vacancies. Vacancies closer to job seekers' home yield higher utility. Our theoretical model is similar to the one by Manning and Petrongolo (2017), where agents must choose a set of places to apply to, and borrows elements from Albrecht, Gautier, and Vroman (2006), and Galenianos and Kircher (2009).

Each firm has one vacancy. The location of vacancies is exogenous and fixed. While this assumption may seem restrictive, geographic mismatch is going to be even smaller if both vacancies and workers are allowed to relocate in order to improve their chances of matching. All workers and all firms are identical, risk neutral, and they produce one unit of output when matched and zero otherwise. The utility of an employed worker is defined below, and an unmatched worker has a utility of zero. Workers observe all vacancies. Workers and vacancies are spread across S locations. Denote by $i(u)$ and $j(v)$ the geographic units where unemployed worker u and vacancy v are respectively located. Each location k has V_k vacancies and U_k unemployed workers. A worker's strategy is a set of $\overline{a}$ vacancies that s/he applies to. The timing of the game is the following:

- Job seekers apply to vacancies: each job seeker sends $\overline{a}$ applications.
- Firms gather the applications they receive: each application has a probability q to be valid in the sense that the applicant will produce positive output if hired. Here, q is a scale parameter: it helps us calibrate the model by capturing the fact that the matching rate in the labor market is lower than what can be predicted on the basis of the number of applications that firms receive. For the sake of tractability, q is assumed to be constant and is not allowed to depend on distance or occupations.[13]
- Firms can only make one offer. If a vacancy has more than one valid application, the firm randomly picks the job seeker to whom it makes an offer.
- Offers are sent to job seekers.
- Job seekers can only accept one job offer. If a job seeker has received more than one offer, he accepts the offer that generates the highest utility.
- Matches are realized. If a firm's chosen applicant rejects the job offer, the firm remains unmatched.

The application of a worker u to a vacancy v provides the worker utility $w_{uv} = f(d_{i(u)j(v)})\varepsilon_{uv}$, the product of a deterministic decreasing function f of

[13] In an extension (online Appendix), we allow q to depend on the previous occupation of the job seeker in a simplified version of the model. We find that doing so slightly lowers mismatch.

the geographic distance $d_{i(u)j(v)}$ between the job seeker and the vacancy, and an idiosyncratic term ε_{uv} that is job-worker pair specific.

Assume that ε is uncorrelated across job seekers and vacancies. This idiosyncratic term ε explains why job seekers do not only apply within their own zip code: some very desirable vacancies (high ε) are located in other zip codes and job seekers have to trade off distance with ε. Furthermore, ε allows for workers in a given location i to have different preferences over vacancy locations j. Finally, ε allows for unobserved job heterogeneity within a location from the point of view of each specific job seeker.

We assume that the probability π_{uv} that a worker u gets an offer for vacancy v conditional on applying only depends on the location of the vacancy: $\pi_{uv} = \pi_{j(v)}$. This assumption, which is crucial for the tractability of the model, is also present in Manning and Petrongolo (2017). In an extension using a simplified mismatch index (online Appendix), we allow the probability of getting an offer to depend on the distance between the applicant and the job. As is intuitive, if the distance between the applicant and the job lowers the offer probability, mismatch increases; yet the increase is modest for realistic assumptions.

We now discuss job seekers' optimal strategy. If vacancies closer from a worker's residence do not have a systematically lower probability of yielding an offer, a worker's optimal strategy is to apply to the $\overline{a}$ vacancies with the highest expected utility.[14] This assumption seems like a reasonable approximation: distance to workers' residence and the probability of getting an offer from a vacancy cannot be systematically negatively correlated because workers are geographically dispersed.[15]

Given job seekers' optimal strategy, we derive p_{ij} the probability for a job seeker in i to apply to a vacancy in j. A job seeker u applies to the $\overline{a}$ vacancies with the highest expected utilities $\pi_{j(v)} f\left(d_{i(u)j(v)}\right)\varepsilon_{uv}$. Assuming that ε has a Pareto distribution[16] of parameter α, p_{ij} is proportional to $\pi_j^{\alpha} f^{\alpha}(d_{ij})$. Given that the total number of applications per job seeker is equal to $\overline{a}$, and denoting $g(d_{ij}) = f^{\alpha}(d_{ij})$, we obtain

$$p_{ij} = \overline{a}\frac{\pi_j^{\alpha} g(d_{ij})}{\sum_{\ell} \pi_{\ell}^{\alpha} g(d_{i\ell}) V_{\ell}}, \quad \forall i,j. \tag{3}$$

The probability of applying p_{ij} increases in the probability of getting an offer from a vacancy in j, π_j, and decreases with distance d_{ij} between the job seeker and the vacancy, according to the distaste for distance function g.

The probability that job seekers match depends on the probability of getting an offer from vacancies in each j where they applied, π_j. To derive π_j, we first need to determine how many valid applications a vacancy receives. The total number of applications received by a vacancy located in j from job seekers located in i is distributed as a Poisson$(p_{ij}U_i)$. Summing applications coming from all origins and keeping only the valid ones (probability q), the distribution of the number of valid

[14] Assuming that a worker can receive at most one offer, as in Manning and Petrongolo (2017), leads to the same optimal strategy.

[15] For a more in-depth discussion of these assumptions, see the online Appendix.

[16] This assumption is also present in Manning and Petrongolo (2017).

applications received by a vacancy in j is a Poisson(qr_j), where $r_j = \sum_k p_{kj} U_k$ is the expected number of applications received by a vacancy in j.

From the point of view of job seekers, the probability π_j that an application generates an offer is the probability that their application is valid (q), and that it is picked out by the firm among all other valid applications that the vacancy has received:[17]

$$\pi_j = q\mathcal{R}(qr_j), \tag{4}$$

where $\mathcal{R}(x) = [1 - \exp(-x)]/x$. Combining equations (3) and (4) and eliminating p and q, we obtain

$$\pi_j = q\mathcal{R}\left(\pi_j^\alpha q\overline{a}\sum_k \frac{g(d_{kj})\,U_k}{\sum_\ell \pi_\ell^\alpha g(d_{k\ell})\,V_\ell}\right). \tag{5}$$

The total number of matches can be expressed as the number of job seekers multiplied by the probability that each job seeker forms a match. The probability that a job seeker matches depends on the number of offers received by a job seeker in i from vacancies in each location j, which is distributed as Poisson$\left(\pi_j p_{ij} V_j\right)$. The total number of offers received by this job seeker in i from all locations is thus distributed as $Y_i \sim$ Poisson$\left(\sum_\ell \pi_\ell p_{i\ell} V_\ell\right)$. A job seeker in k will match if and only if s/he receives at least one offer, which is one minus the probability of getting zero offers, i.e., $1 - \exp\left(-\sum_\ell p_{k\ell}\pi_\ell V_\ell\right)$. Using equation (3) to substitute $p_{k\ell}$ by its expression, the total number of matches M is

$$M = \sum_k U_k\left[1 - \exp\left(-\overline{a}\frac{\sum_\ell \pi_\ell^{1+\alpha} g(d_{k\ell})\,V_\ell}{\sum_\ell \pi_\ell^\alpha g(d_{k\ell})\,V_\ell}\right)\right]. \tag{6}$$

In a nutshell, the total number of matches M is equal to the sum of job seekers weighted by the probability that each job seeker gets at least one offer. In turn, the probability of getting at least one offer (in brackets) depends on the number of vacancies weighted by the probability that a vacancy yields an offer (π_ℓ) and a decreasing function of the distance from the job seeker $(g(d_{k\ell}))$.

How can we determine the total number of hires M given the parameters $\overline{a}$, α, q, $g(\cdot)$, as well as vectors U and V? Once π is known, it is straightforward to find the total number of hires using equation (6). However, π is difficult to pin down because π is a nonlinear function of itself: equation (5) defines a system of S equations where the S π_j are the unknowns, and $\overline{a}$, α, q, $g(\cdot)$, U, and V are the parameters. We do not have a proof for the existence or uniqueness of a solution vector π. However, we find numerically that the expression for π_j in equation (5) defines a contraction mapping that reaches an equilibrium very fast, for a large range of parameters. Trying several

[17] π_j is equal to q multiplied by the expectation of $1/(X_j + 1)$, where X_j is the expected number of valid applications made by other job seekers to the job, with $X_j \sim$ Poisson(qr_j).

starting points always leads to the same solution, which argues in favor of a unique equilibrium.[18]

B. *Estimation of the Structural Parameters*

We now turn to the issue of estimating the structural parameters of the model: g, the distaste for distance, α, the Pareto parameter for the match-specific utility component ε, and q, the probability of a valid application. We start with describing the parameters we set, and then turn to the estimation of the structural parameters.

First, we set $\overline{a}$ as the average number of applications by job seekers observed in the data.[19] Second, because the number of matches in the model depends on labor market tightness, we must make sure that labor market tightness in our data is representative of the US economy. To do so, we apply a proportionality factor to our vacancies so that the aggregate labor market tightness in our data is the same as in the US economy.[20]

Turning to the estimation of the structural parameters, we need to determine the values of g, α, and q given $\overline{a}$ and vectors U and V. In order to estimate the distaste for distance g, we choose the parametrization $g = \exp(\eta s(d))$, where $s(\cdot)$ is the spline already estimated in the reduced-form equation (1) with only job seeker fixed effects, and η is a scalar to be estimated now. As further discussed below, η allows us to adjust for the potential bias arising from our estimating $s(\cdot)$ with job seeker fixed effects only, rather than both job seeker and vacancies fixed effects. Under this parametrization of g, the estimation of g and α amounts to the estimation of two parameters, η and α.

We now explain how the estimation proceeds in order to determine η, α, and q, given U, V, and $\overline{a}$. For each value of (α, η), q is set so that the average job finding rate predicted by the model matches the national job finding rate computed using the CPS.[21] We estimate α and η by maximum likelihood. For a given value of (α, η), we can use our model to compute p_{ij}, the probability that an individual in i applies to a job in j. The value of p_{ij} is directly related to an observed quantity, the number of applications from i to j, A_{ij}. According to our model, A_{ij} is drawn

[18] We can analytically derive π and a closed form for the mismatch index if we assume that $\alpha = 0$, i.e., job seekers do not take into account other job seekers' applications when deciding where to apply for jobs (online Appendix). This closed form mismatch index yields results that are very similar to our preferred mismatch index and can straightforwardly be used to compute mismatch with other datasets. Another way to get a closed form solution is provided by Manning and Petrongolo (2017), who can prove the existence and uniqueness of the solution in a similar case by assuming that q is small enough that job offers made by employers are always accepted.

[19] As the total number of hires depends on the product $q\overline{a}$ and q is estimated, the mismatch index is actually not sensitive to the value chosen for $\overline{a}$.

[20] For each month of April to June 2012, we compute the monthly tightness by dividing the total number of vacancies (from JOLTS) by the total number of unemployed job seekers (as reported by the Bureau of Labor Statistics based on the Current Population Survey) and take the average as our measure of national labor market tightness. Keeping the geographic distribution fixed, we then inflate the number of vacancies in our data such that the global tightness is equal to the national labor market tightness.

[21] The national job finding rate is computed with the CPS as the number of unemployment to employment transitions in a given month divided by the number of unemployed workers in the previous month. We then compute a target number of hires $\widehat{M}$, equal to the national job finding rate times the number of job seekers in our sample. q is estimated as the quantity minimizing the squared difference between the number of hires predicted by the model M and the target $\widehat{M}$.

from a Poisson distribution of parameter $\lambda_{ij} = U_i V_j p_{ij}$. We need to find the values of α and η such that $\lambda_{ij}(\alpha, \eta)$ is the most likely parameter of the Poisson underlying A_{ij}. Formally, we find α and η by maximizing the quasi-log-likelihood $\mathcal{L}(\alpha, \eta) = \sum_{i,j} A_{ij} \log \lambda_{ij}(\alpha, \eta) - \lambda_{ij}(\alpha, \eta)$.[22]

We now give an intuition for the identification of η and α. We start with the parameter of the distaste for distance η. Consider two distinct zip codes 1 and 2. Intuitively, the farther away 1 and 2 are from each other, the fewer across-zip applications there will be relative to within-zip applications; so this comparison of across- versus within-zip applications allows us to track down the distaste for distance. Based on our model, the number of applications is $\lambda_{ij} = U_i V_j p_{ij}$, so this comparison can be written as

$$\frac{\lambda_{12}\lambda_{21}}{\lambda_{11}\lambda_{22}} = \frac{p_{12}p_{21}}{p_{11}p_{22}} = \frac{g(d_{12})\,g(d_{21})}{g(d_{11})\,g(d_{22})} = \exp(2\eta s(d_{12})).$$

Note that the right-hand side does not depend on the parameter α, and thus we conjecture that η can be identified separately from α. The transformation on λ is equivalent to introducing fixed effects for origin and destination zip codes in a reduced-form context. If $s(\cdot)$ was estimated using a two-way fixed effect on the full sample, we would expect to have $\eta = 1$. Using the $s(\cdot)$ estimated with fixed effects on the job seeker zip code[23], the maximum likelihood estimate of η is 1.0020, so the reduced-form estimate of the distaste for distance is essentially unbiased. Thus, we consider in what follows that $g(d) = \exp(s(d))$.

Now, given η, what is the source of identification for α? We can show that the expected number of applications received by a vacancy in j is

$$r_j = q\bar{a}\mathcal{R}^{\alpha}(qr_j)\sum_k U_k g(d_{kj}).$$

In this expression, $q\bar{a}\sum_k U_k g(d_{kj})$ is the number of valid applications that a given vacancy in j would receive if applicants did not factor in the probability of getting an offer π_j. The term $\mathcal{R}^{\alpha}(qr_j)$ plays the role of a moderating force: if a place attracts more applications r_j, $\mathcal{R}^{\alpha}(qr_j)$ decreases and moderates the increase in r_j. The higher α, the more this force is at play. A higher α lowers the dispersion in r_j, so α can be identified by matching the r_j predicted by the model to the number of applications observed in the data. In our main approach, we estimate α, but we show in the online Appendix that assuming $\alpha = 0$ makes the model far more tractable and barely affects the estimated level of mismatch. This simpler mismatch index can also be used for a number of extensions to our baseline model, which would otherwise not

[22] As can be seen in the online Appendix, the log likelihood has a local maximum in the neighborhood of the optimal values of η and α.

[23] The one-way job seeker zip code fixed effect gives essentially the same estimate for the distaste for distance as the two-way fixed effects. However, the two-way fixed effect estimate is only based on 10 percent of the sample, which is why we use the one-way fixed effect. Either way, the estimates are so close that mismatch is not sensitive to using the one or the other.

TABLE 2—MODEL PARAMETERS

Parameter	Notation	Value	Setting procedure
Job seekers in each zip code	U_i		From data
Vacancies in each zip code	V_j		From data, adjusted so that aggregate tightness matches V_{JOLTS}/U_{CPS}
Average number of applications	$\overline{a}$	12.6	From data
Probability of a valid application	q	0.0290	Estimated to match the national CPS job finding rate
Scaling parameter for the disutility of distance	η	1.0020	Estimated to match the geographic distribution of applications
Pareto parameter for match-specific utility shock	α	0.4629	Estimated to match the geographic distribution of applications

be computationally feasible. Furthermore, the online Appendix shows that mismatch stays of the same order of magnitude for a plausible range of values of α.

Table 2 lists the parameters of the model and the values of the estimated parameters. Note that the probability of a valid application q is quite low because the job finding rate in the CPS is only 18.2 percent, despite the fact that job seekers are sending multiple applications. In order to match the CPS job finding rate, we must assume that most applicants are not qualified for the job.

C. *Mismatch Index*

In this section, we assume that a social planner can move job seekers at no cost to maximize the number of hires. Just as in the standard approach (Section II above), we define mismatch as the difference between the maximum number of hires obtained by the planner (M^*) and the number of hires obtained with the actual allocation of job seekers (M): mismatch is then $1 - M/M^*$. Note that this concept of mismatch has no implications for social welfare defined in a general way. Mismatch only represents a deviation from the objective of maximizing the number of matches and thus minimizing aggregate unemployment.

We want to find the allocation of job seekers that maximizes M. If distance to jobs did not matter to job seekers, we would have a single integrated labor market. The social planner would not have to move job seekers: matches M would be maximized, regardless of the location of job seekers. We thus use the case of no distaste for distance to infer how the social planner can maximize hires.

In the online Appendix, we show that, if job seekers have no distaste for distance, the probability of getting an offer π is equalized across locations. This makes sense because, with no distaste for distance, any job is as good as any other (up to ε, which is i.i.d.). Based on this observation, we conjecture that, to maximize hires, the social planner should reallocate job seekers in order to equalize π across locations.

If π is equal across locations j, the average number of applications r_j received by a vacancy does not depend on its location, as there is a bijective relationship between π_j and r_j. Thus, r_j will be equal to the total number of applications divided by the number of vacancies: $\sum_k p_{kj} U_k = \bar{a}\bar{U}/\bar{V}$, where $\bar{U}$ is the total number of unemployed workers in the economy and $\bar{V}$ is the total number of vacancies. Thus, we can rewrite π as

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

and the total number of matches is

$$M^* = \bar{U}\left[1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\frac{\bar{U}}{\bar{V}}\right)\right)\right]. \tag{7}$$

Interestingly, the number of matches obtained with the allocation that equalizes π across zips is identical to the one obtained with any allocation of job seekers in the case where there is no distaste for distance (online Appendix), which supports our initial conjecture that reallocating job seekers to equalize π maximizes matches.[24]

Our interconnected-markets mismatch index is then defined as one minus the ratio between the number of matches with the actual allocation of job seekers and the maximum number of matches:

$$\mathcal{M}_i = 1 - \sum_k \frac{U_k}{\bar{U}} \frac{1 - \exp\left(-\bar{a}\frac{\sum_\ell \pi_\ell^{1+\alpha} g(d_{k\ell}) V_\ell}{\sum_\ell \pi_\ell^{\alpha} g(d_{k\ell}) V_\ell}\right)}{1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\bar{U}/\bar{V}\right)\right)}. \tag{8}$$

In contrast with our approach, most of the existing literature makes the simplifying assumption that markets are distinct, that is: job seekers can only apply to vacancies within their own unit; job seekers are equally likely to apply to all vacancies within their own unit.

If we assume $g(d_{ii}) = 1$ within the unit and $g(d_{ij}) = 0$ if $i \neq j$, we can build a distinct-market mismatch index equal to

$$\mathcal{M}_d = 1 - \sum_k \frac{U_k}{\bar{U}} \frac{1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\, U_k/V_k\right)\right)}{1 - \exp\left(-q\bar{a}\mathcal{R}\left(q\bar{a}\bar{U}/\bar{V}\right)\right)}. \tag{9}$$

[24] One can also define the allocation of job seekers that maximizes hires. Denote $\tilde{V}_k = \sum_\ell g(d_{k\ell}) V_\ell$, X the matrix of term $[g(d_{ij})\tilde{V}_i]_{ij}$, and b a vector of ones (of dimension the number of zip codes). The allocation of job seekers such that π is constant across zip codes is equal to $U^* = \frac{\bar{U}}{b'X^{-1}b}X^{-1}b$.

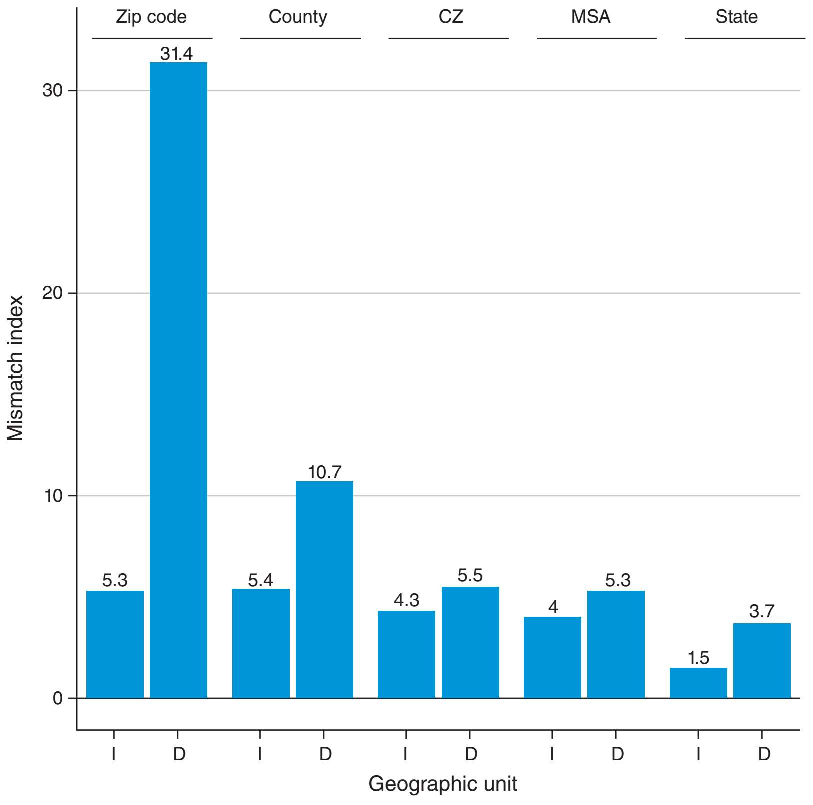

Figure 6. Mismatch Unemployment: Interconnected and Distinct Markets

Source: CareerBuilder.com

D. *Geographic Mismatch: Results*

Mismatch is most accurately captured by our mismatch index $\mathcal{M}_i$ (equation (8)) at the zip code level, because it allows for detailed geography and for interconnected labor markets. Using our preferred measure of mismatch, we find that geographic mismatch is very small: 5.3 percent of hires are lost due to the misallocation of job seekers (Figure 6, interconnected markets).

Our mismatch estimate of 5.3 percent implies that we could reduce US aggregate unemployment by approximately 5.3 percent if we reallocated job seekers to maximize hires.[25] Aggregating the data to the MSA, CZ, county,[26] or zip code level

[25] At the steady state, the unemployment rate is given by $u = \mu_u/(\mu_e + \mu_u)$ with μ_u the entry rate and μ_e the exit rate to/from unemployment. If μ_u is fixed and μ_e increases to μ_e^* when we go to the hires-maximizing allocation, $\mathcal{M} = 1 - \mu_e/\mu_e^*$. The decrease in the steady-state unemployment rate is equal to $(u - u^*)/u = (\mu_e^* - \mu_e)/(\mu_e^* + \mu_u)$. Given that $\mu_e \gg \mu_u$, $(u - u^*)/u \simeq \mathcal{M}$.

[26] The level of mismatch at the county level is slightly higher than at the zip code level. Indeed, unlike the Cobb-Douglas mismatch index, our mismatch index does not monotonically decline when data is more aggregated. When we aggregate the data at the county level, we place all job seekers and all jobs in the middle of the county. Such

consistently yields a mismatch close to 5 percent.[27] It is only if we aggregate the data at the state level that mismatch is markedly smaller. Thus, for a broad range of aggregation levels, mismatch is stable around 5 percent.[28]

The fact that geographic mismatch is low may seem surprising given the differences in unemployment rates across US states. In our model, dispersion in unemployment across states is due to dispersion in labor market tightness. Crucially, our matching function implies a small impact of dispersion in tightness on aggregate unemployment, a feature that is shared by the Cobb-Douglas mismatch index at the state level. Therefore, high dispersion in unemployment rates across states is compatible with low geographic mismatch because such dispersion does not result in large losses in the aggregate number of matches.

A recent literature has attempted to isolate the determinants of workers' location decisions (e.g., Diamond 2016) and to explore the sources of differences in unemployment rates across states (e.g., Herz and van Rens 2015, Amior and Manning 2015). Our results suggest that these differences in unemployment rates across locations do not matter much for aggregate unemployment: whatever its sources, unemployment dispersion accounts for a very limited amount of aggregate unemployment.

The Cobb-Douglas index $\mathcal{M}_{CD}$ (Figure 5) using county level data yields a level of mismatch that is similar to our preferred measure based on zip code data. However, $\mathcal{M}_{CD}$ grossly overestimates mismatch based on zip code data, and underestimates mismatch based on CZ or MSA data. These differences arise both because our model uses a different matching function, and because we allow for applications across geographic units.

To understand the independent role of allowing for applications across geographic areas, we recalculate mismatch with our model, but prevent job seekers from applying across geographic areas ($\mathcal{M}_d$, equation (9)). The distinct markets mismatch index $\mathcal{M}_d$ (Figure 6) yields similar results to those arising from the Cobb-Douglas mismatch index $\mathcal{M}_{CD}$. Just like $\mathcal{M}_{CD}$, $\mathcal{M}_d$ is sensitive to the size of the area where job seekers are assumed to look for jobs, with larger areas yielding smaller levels of mismatch. This similarity suggests that the discrepancy between $\mathcal{M}_i$ and the Cobb-Douglas index is mostly due to the fact that our index accounts for across markets applications rather than to the functional form of the matching function.

Overall, using a search and matching model that fully takes into account the geography of job search and data at the zip code level, we find that eliminating geographic mismatch would reduce US aggregate unemployment by at most 5.3 percent.

an aggregation procedure puts job seekers closer to jobs in their own county but farther away from jobs in other counties, leading to a negative net effect on the number of matches.

[27] When the model is estimated at the zip or county level, we consider that the internal distance (within the same zip or same county) is 0. When the model is estimated at coarser levels (CZ, MSA, state), we follow the trade literature dealing with the estimation of gravity models and introduce the internal distance defined as two-thirds of the square root of the area of the unit divided by π (Head and Mayer 2004).

[28] In a robustness test, we recalculated mismatch at the MSA and CZ levels including vacancies for which we have the city but not the zip code (i.e., essentially all vacancies). The resulting mismatches do not change much: 5.28 percent at the MSA level and 5.47 percent at the CZ level.

IV. Robustness and Extensions

A. *Geographic and Occupational Mismatch*

Mismatch unemployment can be the result of a different geographic distribution of job seekers and job vacancies, but it can also result from a different distribution of job seekers and job vacancies across occupations. Moreover, the occupation and spatial dimensions may interact to further increase mismatch. In this subsection, we move beyond purely geographical mismatch, and compute mismatch combining geographic and occupational heterogeneity.

We define a labor market as a location *and* an occupation and calculate mismatch using these two dimensions at the same time. In order to keep computations tractable, we define labor markets as the intersection of SOC-2 occupations and Commuting Zones, obtaining around 10,000 CZ $\times$ SOC-2 labor markets. For job seekers, their occupation is defined as the occupation of their last job on their résumé. Just as we do not assume that job seekers only apply in their home location, we do not assume that job seekers whose last job was in a given occupation will restrict their applications to the same occupation.

Distaste for Geographic and Occupational Distance.—Restricting applications to be within CZ $\times$ SOC2 would be a bad approximation to reality since only 26 percent of applications are within CZ $\times$ SOC2. Therefore, we need to define an application function that depends both on the geographic distance between CZs and on the occupational distance between SOC2s.

To estimate distance between two SOCs, we use factor analysis, an approach common to the existing literature (Poletaev and Robinson 2008). For each 8-digit SOC, there is a vector (defined by ONet) of about 200 elements that represents the knowledge, skills, and abilities associated with the jobs in this occupation. We perform a factor analysis on these vectors to extract the major dimensions of heterogeneity across occupations. We consider the first two factors: the first one corresponds roughly to the level of intellectual knowledge and abilities required for an occupation (high for executives, for instance), while the second corresponds to physical and technical skills (high for construction workers or electricians, for instance). Then, for each two-digit SOC, we take the mean of each factor across the eight-digit SOCs that together constitute the two-digit SOC.

We estimate a model similar to the one described in equation (1), and we add a dummy for applying to an occupation that is different from the one held in the last job, as well as functions of the two factors estimated above. Specifically, the probability for a job seeker in labor market (i,o) (CZ i and occupation o) to apply for a job in labor market (j,m) (CZ j and occupation m) is

$$\text{(10)} \quad \mu_{(i,o),(j,m)} = U_{i,o}V_{j,m}\exp\Big[\alpha_{i,o} + s(d_{ij}) + \alpha_1 1\{o \neq m\} + \alpha_2\big[(\phi_o - \phi_m)^2 + (\psi_o - \psi_m)^2\big]^{1/2} + \alpha_3(\phi_o - \phi_m) + \alpha_4(\psi_o - \psi_m)\Big]$$

where d_{ij} is the geographic distance between the centroids of the CZs i and j, $\alpha_{i,o}$ is a job seeker CZ $\times$ occupation fixed effect.[29] Note that α_1 estimates a discontinuous preference for one's own SOC2, so we expect $\alpha_1 < 0$; α_2 is the coefficient on the distance between two SOC2 using the two factors ϕ and ψ: we expect $\alpha_2 < 0$; α_3 and α_4 capture the fact that, in each skill dimension, it might be easier to apply to jobs that are less skilled than one's own occupation, so we expect $\alpha_3, \alpha_4 > 0$. In the online Appendix, we show the results of these estimates, and confirm the predictions about the sign of the coefficients. For example, the estimates imply that job seekers are 2.8 times less likely to apply to a SOC2 different from their own, even after accounting for levels of each SOC2 in the two main factors, as well as the differences in factors between the two SOC2. Clearly, job seekers prefer their own SOC2, but this preference is not overwhelming and hence it is necessary to model across-SOC2 applications.

Geographic and Occupational Mismatch: Results.—Plugging the estimates of the distaste for geographic and occupational distance into our mismatch index $\mathcal{M}_i$ and using the same parameter estimation procedure as for our main estimate (i.e., as in Table 2), we find that 6.9 percent of hires are lost due to a combination of geographic and occupational mismatch (Figure 7). Thus, the mismatch index corresponding to the CZ $\times$ SOC-2 labor markets is higher than the one corresponding to interconnected CZ labor markets (4.25 percent), or even distinct CZ labor markets (5.5 percent) (Figure 7).

How does heterogeneity by occupation contribute to mismatch compared to heterogeneity across geography? To shed light on this question, we shut down applications across CZ, across SOC2, or both (Figure 7). If we shut down applications across CZ (distinct CZ) but still allow job seekers to apply across SOC2, mismatch is 8.0 percent. If we instead shut down applications across SOC2 (distinct SOC2) but allow applications across CZ, mismatch is 14.6 percent, which is twice as high as the level of geographic mismatch estimated when allowing applications across occupations. Finally, in the case where we shut down applications both across CZs and across SOCs, thus assuming distinct CZ $\times$ SOC2 markets, mismatch is 17.2 percent.[30] We conclude that mismatch is more severely overestimated by not allowing for applications across SOC-2 than by not allowing for applications across CZ.

Mismatch by Detailed SOC Code and by Education.—Since in the previous section we used a relatively coarse grouping of occupations, this could underestimate mismatch. Here we compute mismatch for the most common eight-digit SOC among job seekers (customer service representatives) and the most common eight-digit SOC among vacancies (registered nurse). To compute mismatch for customer service representatives, we only keep job seekers and jobs in this occupation; the same applies

[29] We also estimated the model with job CZ $\times$ occupation fixed effects, and estimates are very similar as shown in the online Appendix.

[30] There is an interaction effect whereby not allowing for geographically interconnected markets yields a larger increase in mismatch if we also do not allow for interconnected occupations: 14.6 percent to 17.2 percent versus 6.9 percent to 8 percent in the case of interconnected occupations.

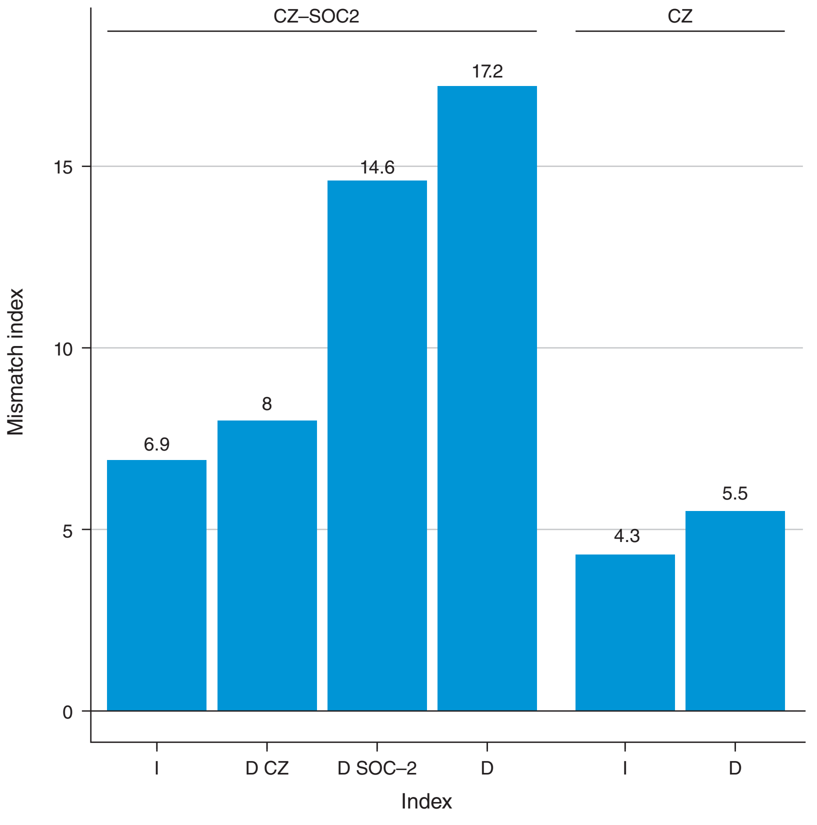

Figure 7. Mismatch When Markets Are Defined as Interaction of Geographic and Occupational Units (CZ × SOC-2) and as CZ

Source: CareerBuilder.com

to mismatch for registered nurses. We estimate mismatch with occupation-specific distaste for distance (Figure 4), and leave all other parameters as in the baseline.

For registered nurses, 5.1 percent of hires are lost due to mismatch. For customer service representatives, mismatch is much higher at 9.3 percent. The higher mismatch for customer service representatives is mostly due to their worse distribution across the territory (farther away from jobs) rather than to their greater distaste for distance. Indeed, if we assume that customer service representatives have the same distaste for distance as registered nurses, mismatch for customer service representatives is still 8.6 percent. Note that these mismatch indices are overestimated because we computed them under the assumption that job seekers only apply to jobs in their past six-digit occupation. The overall conclusion is that mismatch can vary considerably across occupations, but it stays relatively small even for occupations that are more prone to mismatch.

We also compute mismatch unemployment by level of education (high school, associate's degree, bachelor's degree, and above), assuming that job seekers only apply to jobs in their own education. We find that mismatch unemployment decreases with the level of education consistent with more educated workers being more willing to apply far away from home at long distances (Figure 3). Yet, mismatch is never much higher than 5 percent (online Appendix).

Overall, we find that occupational mismatch is considerably overestimated if we ignore applications across occupations. Eliminating *both* geographical (CZ) and occupational (SOC-2) mismatch would reduce US aggregate unemployment by only 6.9 percent, and not 17.2 percent as would be the case if job seekers did not apply across occupations and CZs.

B. *Mismatch for Various Distastes for Distance*

People who use CareerBuilder.com for job search may have lower distaste for distance. So, would mismatch increase a lot if distaste for distance were greater? We compute mismatch at the zip code level and for different distastes for distance. We rely on our baseline estimate of the distaste for distance and parametrize it with ξ: $g = \exp(\xi s(d))$, where $s(\cdot)$ is the spline estimated in the reduced-form equation (1). We let the parameter ξ vary between 0 and 10 in increments of 0.5.

An increase in the distaste for distance starting from our baseline of $\xi = 1$ barely increases geographic mismatch (Figure 8, actual allocation). If we multiply the distaste for distance by two, mismatch increases to 6 percent, and even if we multiply the distaste for distance by 5, mismatch is still only 7.8 percent. Thus, even if our data underestimates the distaste for distance compared to a representative sample, this barely affects the level of geographic mismatch.

How high would mismatch be if American job seekers had the same distaste for distance as the British job seekers?

American job seekers are eight times[31] more willing to apply to vacancies far away from home than the British job seekers[32] studied by Manning and Petrongolo (2017). When we plug the British distaste for distance in our model, we find that the US mismatch almost doubles, at 10.8 percent.

Despite the fact that British job seekers have a greater distaste for distance than American job seekers, Manning and Petrongolo (2017) found that place-based policies are rather ineffective. Indeed, job seekers from other areas apply to newly created jobs in target areas, so the positive employment effect for target area residents is muted. Because job seekers in the United States apply much farther away from their home area than in the United Kingdom, the impact of place-based policies in the contemporary United States is likely to be even more muted.

We have just seen that the level of mismatch is not very sensitive to the distaste for distance, which suggests that job seekers are fairly close to jobs already. On the other hand, if job seekers were allocated uniformly across space (i.e., the number of

[31] Job 0 is preferred to job 1 if and only if $g(d_1)\varepsilon_1 > g(d_0)\varepsilon_0$. Because ε_0 and ε_1 are Pareto, the probability to prefer a job at distance d rather than the one at distance 0 is after some algebra $\exp(s(d_1))/2$. In our case, this amounts to $\exp(-0.0471397 \times 6.2)/2 = 0.37$; with the estimates found in Manning and Petrongolo (2017), $\exp(-0.3 \times 10) = 5$ percent.

[32] This difference may reflect differences between the US and UK labor markets. It might also be influenced by the methodology used: Manning and Petrongolo (2017) infers the distaste for distance parameter from the estimation of a search-and-matching model but cannot directly observe job seekers' application behavior.

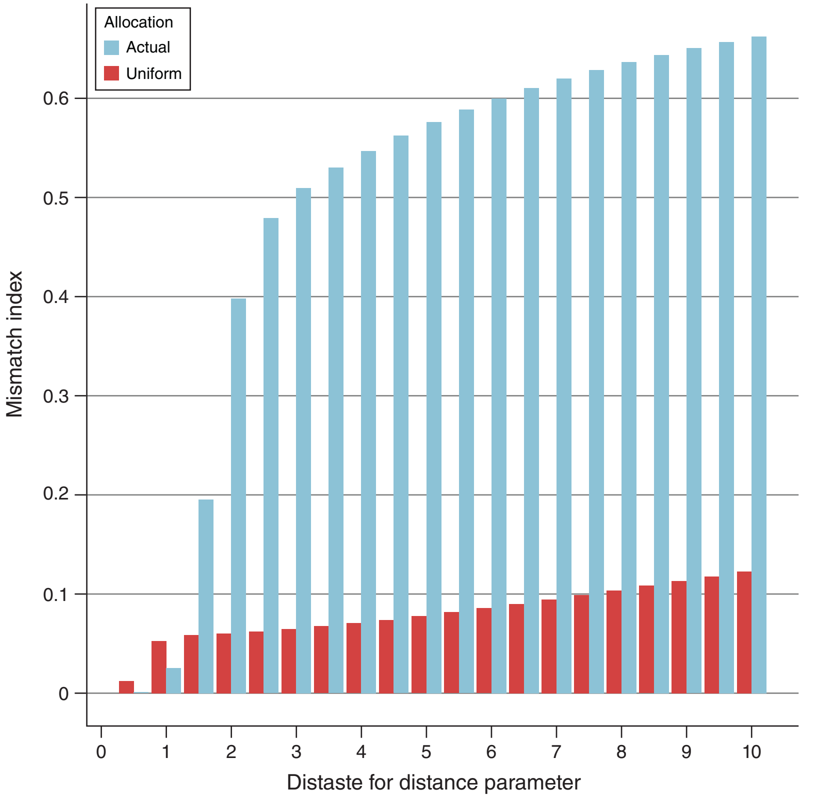

Figure 8. Robustness to Various Distaste-for-Distance Parameters, Actual and Uniform Allocations

Source: CareerBuilder.com

job seekers in a zip code depends on the zip code area), greater distaste for distance would dramatically increase geographic mismatch. For most values of the distaste for distance, mismatch is much higher with the uniform than with the actual allocation (Figure 8, compare dark and light bars).[33] With a uniform allocation and the British distaste for distance, mismatch would be as high as 63.7 percent! These results suggest that, with the actual allocation of job seekers, increasing the distaste for distance has little impact on mismatch because job seekers already live pretty close to vacancies on average.

Based on this analysis, geographic mismatch is low because distaste for distance is low enough, and job seekers are already fairly close to vacancies. In a dynamic framework, low distaste for distance can explain why job seekers are relatively well

[33] When the distaste for distance is very low, the uniform allocation of job seekers yields a lower level of geographic mismatch than the actual allocation of job seekers. This is most likely due to the fact that job seekers are overly concentrated close to big job centers, which reduces their job finding probability and makes them miss out on vacancies that are farther afield. Thus, the job finding rate predicted by our model is lower close to business centers (defined as the zip code in each state with the highest number of vacancies). For low distaste for distance, the uniform allocation fixes this issue by placing job seekers farther away from business centers.

allocated across space: over time, job seekers relocate to follow vacancies so that, at any given point in time, job seekers live close to vacancies on average.

V. Conclusion

In this paper, we have used a novel dataset from CareerBuilder.com to document how far job seekers are willing to apply to jobs and, based on this evidence, we have measured the degree of geographic mismatch. Our measure of geographic mismatch is based on a search and matching model of the labor market in which job seekers strategically choose where to send their applications. Quantitatively, we find that US aggregate unemployment would be reduced by at most 5.3 percent if job seekers were reallocated so as to maximize hires. Therefore, geographic mismatch is a minor driver of US unemployment.

Low mismatch can be explained by job seekers' high enough willingness to apply far away from home combined with the fact that the typical job seeker does not live very far away from jobs. We also extend our model to measure geographic and occupational mismatch taken together. Adding the occupational dimension (two-digit SOC codes) naturally increases mismatch. Yet, geographic and occupational mismatch remains low (6.9 percent) as long as job seekers are allowed to apply across occupations.

Overall, we find that geographic mismatch is a minor cause of unemployment at the macro level. Thus, policies that attempt to combat geographic mismatch by reducing barriers to worker mobility or moving job seekers and jobs closer to each other are likely to have a limited effect on aggregate unemployment.

REFERENCES

Albrecht, James W., Pieter A. Gautier, and Susan Vroman. 2006. "Equilibrium Directed Search with Multiple Applications." *Review of Economic Studies* 73 (4): 869–91.

Altonji, Joseph G., and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives." In *Immigration, Trade and the Labor Market*, edited by John M. Abowd and Richard B. Freeman, 201–34. Chicago: University of Chicago Press.

Amior, Michael, and Alan Manning. 2015. "The Persistence of Local Joblessness." Centre for Economic Performance (CEP) Discussion Paper 1357.

Åslund, Olof, John Östh, and Yves Zenou. 2010. "How important is access to jobs? Old question—improved answer." *Journal of Economic Geography* 10 (3): 389–422.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The Geography of Trade and Technology Shocks in the United States." *American Economic Review* 103 (3): 220–25.

Autor, David H., David Dorn, and Gordon H. Hanson. 2015. "Untangling Trade and Technology: Evidence from Local Labor Markets." *Economic Journal* 125 (584): 621–46.

Barlevy, Gadi. 2011. "Evaluating the Role of Labor Market Mismatch in Rising Unemployment." *Economic Perspectives* 35 (3): 82–96.

Barnichon, Regis, and Andrew Figura. 2015. "Labor Market Heterogeneity and the Aggregate Matching Function." *American Economic Journal: Macroeconomics* 7 (4): 222–49.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 22 (1): 1–75.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and John van Reenen. 2004. "Evaluating the Employment Impact of a Mandatory Job Search Program." *Journal of the European Economic Association* 2 (4): 569–606.

Borjas, George J. 2003. "The Labor Demand Curve *is* Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118 (4): 1335–74.

Borjas, George J., Richard B. Freeman, and Lawrence F. Katz. 1996. "Searching for the Effect of Immigration on the Labor Market." *American Economic Review* 86 (2): 246–51.

Borjas, George J., Richard B. Freeman, and Lawrence F. Katz. 1997."How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* 27 (1): 1–67.

Bound, John, and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics* 18 (1): 20–54.

Boustan, Leah Platt, and Robert A. Margo. 2009. "Race, segregation, and postal employment: New evidence on spatial mismatch." *Journal of Urban Economics* 65 (1): 1–10.

Busso, Matias, Jesse Gregory, and Patrick Kline. 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review* 103 (2): 897–947.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43 (2): 245–57.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19 (1): 22–64.

Card, David. 2005. "Is the New Immigration Really so Bad?" *Economic Journal* 115 (507): F300–F323.

Card, David, and John DiNardo. 2000. "Do Immigrant Inflows Lead to Native Outflows?" *American Economic Review* 90 (2): 360–67.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* 128 (2): 531–80.

Davidson, Carl, and Stephen A. Woodbury. 1993. "The Displacement Effect of Reemployment Bonus Programs." *Journal of Labor Economics* 11 (4): 575–605.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2012."Recruiting Intensity during and after the Great Recession: National and Industry Evidence." *American Economic Review* 102 (3): 584–88.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524.

Fan, Yingling. 2012. "The Planners' War against Spatial Mismatch: Lessons Learned and Ways Forward." *Journal of Planning Literature* 27 (2): 153–69.

Farber, Henry S. 2012. "Unemployment in the Great Recession: Did the Housing Market Crisis Prevent the Unemployed from Moving to Take Jobs?" *American Economic Review* 102 (3): 520–25.

Ferracci, Marc, Grégory Jolivet, and Gerard J. van den Berg. 2014."Evidence of Treatment Spillovers Within Markets." *Review of Economics and Statistics* 96 (5): 812–23.

Friedberg, Rachel M., and Jennifer Hunt. 1995. "The Impact of Immigrants on Host Country Wages, Employment and Growth." *Journal of Economic Perspectives* 9 (2): 23–44.

Galenianos, Manolis, and Philipp Kircher. 2009. "Directed search with multiple job applications." *Journal of Economic Theory* 144 (2): 445–71.

Gautier, Pieter, Paul Muller, Bas van der Klaauw, Michael Rosholm, and Michael Svarer. 2012."Estimating Equilibrium Effects of Job Search Assistance." Institute for the Study of Labor (IZA) Discussion Paper 6748.

Greenwood, Michael J., Gary L. Hunt, and John M. McDowell. 1986. "Migration and Employment Change: Empirical Evidence on the Spatial and Temporal Dimensions of the Linkage." *Journal of Regional Science* 26 (2): 223–34.

Guglielminetti, Elisa, Rafael Lalive, Philippe Ruh, and Etienne Wasmer. 2015. "Spatial search strategies of job seekers and the role of unemployment insurance." https://www.unine.ch/files/live/sites/irene/files/shared/documents/Recherche%20et%20mandats/S%C3%A9minaires%20en%20%C3%A9conomie/2016-2017/Commuting_EG_RL_PR_EW.pdf.

Guimarães, Paulo, and Pedro Portugal. 2010. "A simple feasible procedure to fit models with high-dimensional fixed effects." *Stata Journal* 10 (4): 628–49.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52 (4): 909–38.

Head, Keith, and Thierry Mayer. 2004. "The Empirics of Agglomeration and Trade." In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2609–69. Amsterdam: North-Holland.

Hellerstein, Judith K., David Neumark, and Melissa McInerney. 2008. "Spatial mismatch or racial mismatch?" *Journal of Urban Economics* 64 (2): 464–79.

Herz, Benedikt, and Thijs van Rens. 2015. "Accounting for Mismatch Unemployment." Institute for the Study of Labor (IZA) Discussion Paper 8884.

Ihlanfeldt, Keith R., and David L. Sjoquist. 1998. "The spatial mismatch hypothesis: A review of recent studies and their implications for welfare reform." *Housing Policy Debate* 9 (4): 849–92.

Jackman, R., R. Layard, and C. Pissarides. 1989. "On Vacancies." *Oxford Bulletin of Economics and Statistics* 51 (4): 377–94.

Jackman, R., and S. Roper. 1987. "Structural Unemployment." *Oxford Bulletin of Economics and Statistics* 49 (1): 9–36.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6: 629–62.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *Quarterly Journal of Economics* 128 (3): 1123–67.

Lazear, Edward P., and James R. Spletzer. 2012. "Hiring, Churn and the Business Cycle." *American Economic Review* 102 (3): 575–79.

Manning, Alan, and Barbara Petrongolo. 2017. "How Local Are Labor Markets? Evidence from a Spatial Job Search Model." *American Economic Review* 107 (10): 2877–2907.

Marinescu, Ioana, and Roland Rathelot. 2018. "Mismatch Unemployment and the Geography of Job Search: Dataset." *American Economic Journal: Macroeconomics*. http://doi.org/10.1257/mac.20160312.

Marinescu, Ioana, and Ronald Wolthoff. 2015. "Opening the Black Box of the Matching Function: The Power of Words." Institute for the Study of Labor (IZA) Discussion Paper 9071.

McKenzie, Brian. 2013. "County-to-County Commuting Flows: 2006–10." https://www.census.gov/library/working-papers/2013/acs/2013-McKenzie.html (accessed February 20, 2015).

Molloy, Raven, Christopher L. Smith, and Abigail K. Wozniak. 2011."Internal Migration in the United States." *Journal of Economic Perspectives* 25 (3): 173–96.

Neumark, David, and Jed Kolko. 2010. "Do enterprise zones create jobs? Evidence from California's enterprise zone program." *Journal of Urban Economics* 68 (1): 1–19.

Neumark, David, and Helen Simpson. 2015. "Place-Based Policies." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 1197–1287. Amsterdam: Elsevier.

Nickell, Stephen. 1982. "The Determinants of Equilibrium Unemployment in Britain." *Economic Journal* 92 (367): 555–75.

Notowidigdo, Matthew J. 2011. "The Incidence of Local Labor Demand Shocks." National Bureau of Economic Research (NBER) Working Paper 17167.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2006. "The economic value of cultural diversity: Evidence from US cities." *Journal of Economic Geography* 6 (1): 9–44.

Patacchini, Eleonora, and Yves Zenou. 2005. "Spatial mismatch, transport mode and search decisions in England." *Journal of Urban Economics* 58 (1): 62–90.

Poletaev, Maxim, and Chris Robinson. 2008. "Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000." *Journal of Labor Economics* 26 (3): 387–420.

Rupert, Peter, and Etienne Wasmer. 2012. "Housing and the labor market: Time to move and aggregate unemployment." *Journal of Monetary Economics* 59 (1): 24–36.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2014. "Mismatch Unemployment." *American Economic Review* 104 (11): 3529–64.

Valletta, Robert G. 2013. "House lock and structural unemployment." *Labour Economics* 25: 86–97.

Veracierto, Marcelo. 2011. "Worker Flows and Matching Efficiency." *Economic Perspectives* 35 (4): 147–69.

Wickham, Hadley. 2009. *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wozniak, Abigail. 2010. "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources* 45 (4): 944–70.

Yagan, Danny. 2017. "Employment Hysteresis from the Great Recession." National Bureau of Economic Research (NBER) Working Paper 23844.