Mismatch Unemployment[*][†]

Ayşegül Şahin
Federal Reserve Bank of New York

Joseph Song
Columbia University

Giorgio Topa
Federal Reserve Bank of New York and IZA

Giovanni L. Violante
New York University, CEPR, and NBER

FIRST VERSION: NOVEMBER 2010 - THIS REVISION: SEPTEMBER 2012

Abstract

We develop a framework where mismatch between vacancies and job seekers across sectors translates into higher unemployment by lowering the aggregate job-finding rate. We use this framework to measure the contribution of mismatch to the recent rise in U.S. unemployment by exploiting two sources of cross-sectional data on vacancies: JOLTS and HWOL (a new database covering the universe of online U.S. job advertisements). Mismatch across industries and occupations explains at most 1/3 of the total observed increase in the unemployment rate. Geographical mismatch plays no apparent role. Occupational mismatch has become especially more severe for college graduates, and in the West of the United States.

[*]We are especially grateful to June Shelp, at The Conference Board, for her help with the HWOL data. The opinions expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

[†]This draft is a revised version of NBER WP 18265 which includes (i) a more thorough comparison between HWOL and JOLTS data, (ii) improved estimations of matching functions, and (iii) some additional results.

1 Introduction

The U.S. unemployment rate rose from an average value of 4.6% in 2006 to its peak of 10% in October 2009, as the economy experienced the deepest downturn in the postwar period. Two years after its peak, the unemployment rate still hovered above 8%. This persistently high rate has sparked a vibrant debate among economists and policymakers. The main point of contention is the nature of these sluggish dynamics and, therefore, the appropriate policy response.

A deeper look at worker flows into and out of unemployment shows that, while the inflow rate has returned to its pre-recession level, the job-finding rate is still half of what it was in 2006. Any credible explanation for the recent dynamics in unemployment must therefore operate through a long-lasting decline in the outflow rate. One such theory is that the recession has produced a severe sectoral mismatch between vacant jobs and unemployed workers: idle workers are seeking employment in sectors (occupations, industries, locations) different from those where the available jobs are. Such misalignment between the distribution of vacancies and unemployment across sectors of the economy would lower the aggregate job-finding rate.

The mismatch hypothesis is qualitatively consistent with three features of the Great Recession. First, over the past three years the U.S. Beveridge curve (i.e., the empirical relationship between aggregate unemployment and aggregate vacancies) has displayed a marked rightward movement indicating that, for a given level of vacancies, the current level of unemployment is higher than that implied by the historical data.[1] Put differently, aggregate matching efficiency has declined.[2] Second, around half of the job losses in this downturn were concentrated in construction and manufacturing.[3] To the extent that the unemployed in these battered sectors do not search for (or are not hired in) jobs in the sectors which largely weathered the storm (e.g., health care), mismatch would arise across occupations and industries. Third,

[1]See, among others, Elsby, Hobijn, and Şahin (2010), Hall (2010), Daly, Hobijn, Şahin, and Valletta (2011), Barlevy (2011), and Veracierto (2011). According to these studies, at the current level of vacancies, the pre-recession U.S. unemployment-vacancies relationship predicts an unemployment rate between 2 and 3 percentage points lower than its current value.

[2]According to Barlevy (2011) and Veracierto (2011), the size of this drop from its pre-recession level is between 15% and 25%, depending on the exact methodology used in the calculation.

[3]According to the Current Employment Statistics (CES), also known as the establishment survey, payroll employment declined by 7.4 million during the recession and construction and manufacturing sectors accounted for 54% of this decline.

house prices experienced a sharp fall, especially in certain regions (see e.g., Mian and Sufi, 2011). Homeowners who expect their local housing markets to recover may choose to forego job opportunities in other locations to avoid large capital losses from selling their house. Under this "house-lock"conjecture, mismatch between job opportunities and job seekers would arise mostly across locations.

In this paper, we develop a theoretical framework to conceptualize the notion of mismatch unemployment, and use this framework to measure how much of the recent rise in the U.S. unemployment rate is attributable to mismatch. We envision the economy as comprising a large number of distinct labor markets or sectors (e.g., segmented by industry, occupation, geography, or a combination of these attributes). Each labor market is frictional, i.e., its hiring process is governed by a matching function. To assess the existence of mismatch in the data, we ask whether, given the observed distribution of productive efficiency, matching efficiency, and vacancies across labor markets in the economy, unemployed workers are "misallocated,"i.e., they search in the wrong sectors. Answering this question requires comparing the actual allocation of unemployed workers across sectors to an ideal allocation. The ideal allocation that we choose as our benchmark is the one that would be selected by *a planner who faces no impediment in moving idle labor across sectors except for the within-market matching friction*. We show that optimality for this planner dictates that (productive and matching) efficiency-weighted vacancy-unemployment ratios be equated across sectors. By manipulating the planner's optimality condition, we construct a mismatch index that measures the fraction of hires lost every period because of misallocation of job seekers. Through this index, we can quantify how much lower the unemployment rate would be in the absence of mismatch. The difference between the observed unemployment rate and this counterfactual unemployment rate is *mismatch unemployment*.[4]

Our measurement exercise requires disaggregated data on unemployment and vacancies. The standard micro data sources for unemployment and vacancies are, respectively, the Current Population Survey (CPS) and the Job Openings and Labor Turnover Survey (JOLTS). Unfortunately, JOLTS only allows disaggregation of va-

[4]Our focus is on mismatch unemployment intended as unemployed searching in the "wrong" sector. A separate literature uses the term "mismatch" to denote the existence of employed individuals working on the "wrong" job—meaning a sub-optimal joint distribution of worker skills and firm's capital. See, for example, Eeckhout and Kircher (2011).

cancies by 2-digit industries and very broad geographical area (4 Census regions).[5] In this paper, we introduce a new source of micro data, the Conference Board's Help Wanted OnLine (HWOL) database, designed to collect the universe of unique online job advertisements in the U.S. economy. Through this novel data set, we are able perform our empirical analysis at the 2- and 3-digit occupational level, at a more detailed geographical level (states and counties), and even by defining labor markets as a combination of occupation and location.[6]

Our empirical analysis yields the following main results. We find no significant role for geographical mismatch across U.S. states or counties. Mismatch at the 2-digit industry and 2- and 3-digit occupation level increased markedly during the recession but declined throughout 2010, an indication of a strong cyclical pattern in mismatch. A similar, but milder, hump shape in mismatch is observed around the 2001 recession. With all the caveats associated to a short sample, we do not find evidence of a significant long-run "structural" shift in mismatch after the Great Recession.

We calculate that an additional four percent of monthly hires were lost during the Great Recession because of the misallocation of vacancies and job seekers across occupations and industries. As a result, our counterfactual analysis indicates that mismatch unemployment at the 2-digit industry level can account for 0.75 percentage points out of the 5.4 percentage point total increase in the U.S. unemployment rate from 2006 to October 2009. At the 3-digit occupation level, the contribution of mismatch unemployment rises just beyond 1.5 percentage points. When we compute occupational mismatch separately for different education groups and different Census regions, we find its contribution to the observed increase in the unemployment rate is the largest among college graduates and for the West, and it is the smallest among high-school dropouts and in the North-East.

In an extension of the baseline analysis, we allow the misallocation of unemployed workers across sectors to also affect the vacancy creation decisions of firms: the presence of job-seekers in declining sectors makes it easier to fill jobs in those sectors and, therefore, distorts firms' incentives in the direction of, inefficiently, creating vacancies in the wrong markets. We show that this channel depresses aggregate

[5] Note that industry classification in the JOLTS is slightly different than the 2-digit NAICS classification. See Table B1 for a complete list of industries in the JOLTS.

[6] The HWOL micro data would allow an even more disaggregated analysis. The binding constraint is the small sample size of unemployed workers in the monthly CPS.

vacancy creation relative to the planner's solution, giving a further boost to mismatch unemployment. When this additional force is factored into our counterfactuals, the contribution of mismatch to the observed rise in the unemployment rate grows by a maximum of two thirds of a percentage point. We therefore conclude that, at the analyzed level of disaggregation, mismatch can explain at most 1/3 of the recent rise in the U.S. unemployment rate since 2006.

We now return briefly to the nature of our measurement exercise. Formalizing mismatch unemployment as "distance from a benchmark allocation" follows, in essence, the same insights of the vast literature on misallocation and productivity (Lagos, 2006; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Jones, 2011; Moll, 2011). Our implementation has two distinctive features. First, we do not need to solve for equilibrium allocations (and, hence, make specific assumptions about firms' and workers' behavior, their information set, price determination, etc.) We simply take the empirical joint distribution of unemployment and vacancies across sectors as the equilibrium outcome.[7] Second, we construct the counterfactual distribution (in absence of mismatch) from a simple planner's problem which can be solved analytically.

The key strength of these two features combined is that finer disaggregation in the available micro data poses no threat to the feasibility of the exercise. The approach we propose is robust and easily implementable, even with a high number of labor markets, and multiple sources of heterogeneity, idiosyncratic shocks, and aggregate fluctuations. Our method is also useful in identifying the relative importance of different dimensions (e.g., industry, occupation, geography) of mismatch. The limitation is that one cannot separately quantify the, possibly many, sources of misallocation. This would require specifying and solving a complex structural equilibrium model which, at the level of disaggregation of our analysis, would be computationally unfeasible. Factors explaining the discrepancy between the empirical and planner's distribution of unemployment across sectors include moving (e.g., retraining or migration) costs, relative wage rigidity, or certain government policies that may hamper the reallocation of idle labor from shrinking to expanding sectors. Since moving costs are characteristics of the physical environment that would also

[7]The extension to endogenous vacancy requires a minimal set of, mostly standard, assumptions that are discussed in Section 7.

feature in a planner's problem, while our benchmark planner's allocation is derived under costless between-sector mobility, our calculations on the role of mismatch are an upper bound. In light of this remark, the finding that mismatch is not a chief determinant of the persistently high U.S. unemployment appears even more compelling.

The model underlying our measurement exercise is a multi-sector version of the standard aggregate search/matching model (Pissarides, 2000). Shimer (2007) proposed an alternative environment to measure mismatch between firms and workers across labor markets. The crucial difference between the two models is the notion of a vacancy or, equivalently, at which point of the meeting process vacancies are measured. In the matching model, firms desiring to expand post vacancies: a vacancy is a manifestation of a firm's *effort to hire*. In Shimer's model, firms unsuccessful in meeting workers are left with idle jobs: a vacancy is therefore a manifestation of a firm's *failure to hire*. Both notions are theoretically correct. Since both models are parameterized using the same micro-data on vacancies, the key question is whether existing job-openings data from JOLTS and HWOL are more likely to represent firms' hiring effort or hiring failure. The short duration of job openings in JOLTS (2-4 weeks according to Davis, Faberman, and Haltiwanger, 2010) seems somewhat more consistent with the former view, but better data is needed to shed light on this critical point.

The notion of vacancy we adopt is common to the entire search/matching approach to unemployment. Within this class, the closest paper to ours is Jackman and Roper (1987): in a static matching model with many sectors, they show that distributing unemployment across sectors so that sectoral labor-market tightnesses are equalized maximizes aggregate hires, and they propose the use of mismatch indexes to summarize deviations from this allocation.[8] At that time, economists were strug-

[8]This idea goes back, at least, to Mincer (1966, page 126) who writes: "To detect the existence, degree, and changes in structural unemployment, (U, V) maps may be constructed for disaggregations of the economy in the cross-section, by various categories, such as industry, location, occupation, and any other classification of interest. For example, each location is represented by a point in the (U, V) map, and a scatter diagram showing such information for all labor markets may show a clear positive correlation. This would indicate that unemployment is largely nonstructural with respect to location, that is to say, that adjustments require movements within local areas rather than the more difficult movements between areas. In contrast, a negative relation in the scatter would indicate the presence of a structural problem. The scatters may, of course, show identifiable combinations of patterns. Observations of changes in these cross sectional patterns over time will show rotations and shifts, providing highly suggestive leads for diagnoses of the changing structure of labor supplies and demands."

gling to understand why high unemployment was so persistent in many European countries. Padoa-Schioppa (1991) contains a number of empirical studies for various countries and concludes that mismatch was not an important explanation of the dynamics of European unemployment in the 1980s.[9] Our paper contributes to reviving this old literature by extending it in several directions: (i) we develop a dynamic, stochastic, environment with numerous sources of heterogeneity, (ii) we explain how to construct counterfactual measures of unemployment, absent mismatch, (iii) we incorporate the effect of misallocation on vacancy creation, and (iv) we perform our measurement at a much more disaggregated level, thanks to new micro data.

Beyond the present paper, a small but rapidly growing literature attempts to quantify whether mismatch played a substantive role in the Great Recession. Barnichon and Figura (2011) document that fluctuations in the estimated residuals of the aggregate matching function are negatively correlated with the dispersion of tightness across U.S. labor markets. This empirical finding is in line with our theoretical result showing that our mismatch index acts as a shifter of the aggregate matching function. However, their empirical analysis is mostly disaggregated at the geographical level, a negligible dimension of mismatch according to our calculations. Dickens (2010) and Lazear and Spletzer (2012) study mismatch in the labor market using an alternative index proposed by Mincer (1966). In a previous version of this paper, we also reported results based on this index and argued that is much less useful than the one we propose here because it only quantifies the number of job-seekers searching in the wrong sectors, but not how such misallocation lowers the job-finding rate and raises unemployment. In addition, the analysis in these papers does not allow for heterogeneity in productive and matching efficiency across sectors, a key determinant of the optimal allocation of job-seekers across labor markets. The analysis of Herz and van Rens (2011) is quite complementary to ours: they put more structure on the determination of equilibrium unemployment in order to disentangle various potential sources of mismatch, and as a consequence have stricter data requirements. They conclude that relative wage rigidity (across states and industries) is vastly more

[9]The conjecture was that the oil shocks of the 1970s and the concurrent shift from manufacturing to services induced structural transformations in the labor market that permanently modified the skill and geographical map of labor demand. From the scattered data available at the time, there was also some evidence of shifts in the Beveridge curve for some countries. Subsequent explanations of European unemployment based on the interaction between technological changes and rigid labor market institutions were more successful quantitatively.

important than moving costs as a source of mismatch. In light of their finding, our planner problem may provide a tight upper bound.

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework. Section 3 derives the mismatch indexes and explains how we compute our unemployment rate counterfactuals. Section 4 describes the data. Section 5 performs the empirical analysis. In Section 6 we verify the robustness of our results to measurement errors in unemployment and vacancy counts. Section 7 analyzes the case in which mismatch also affects vacancy creation. Section 8 concludes. Appendix A contains the proofs of our theoretical results and Appendix B contains more detail about the data and our measurement exercise.

2 Environment and planner problem

We begin by describing our economic environment and deriving the planner's optimal allocation rule of unemployed workers across sectors—the crucial building block of our theoretical analysis. Throughout these derivations, we maintain the assumption that the evolution of the vacancy distribution is exogenous. We relax this assumption in Section 7.

2.1 Benchmark environment

Time is discrete. The economy is comprised of a large number I of distinct labor markets (sectors) indexed by i. New production opportunities, corresponding to job vacancies (v_i), arise exogenously across sectors.[10] The economy is populated by a measure one of risk-neutral individuals who can be either employed in sector i (e_i) or unemployed and searching in sector i (u_i). Therefore, $\sum_{i=1}^{I}(e_i+u_i)=1$. On-the-job search is ruled out, and an unemployed worker, in any given period, can search for vacancies in one sector only.

Labor markets are frictional: new matches, or hires, (h_i) between unemployed workers (u_i) and vacancies (v_i) in market i are determined by the matching function

[10]We explain in Section 7 that assuming that vacancies are exogenous is equivalent to a model where the job creation margin is endogenous, and the elasticity of the cost of creating vacancies is infinitely large.

$\Phi \cdot \phi_i \cdot m\,(u_i, v_i)$, with m strictly increasing and strictly concave in both arguments and homogeneous of degree one in (u_i, v_i). The term $\Phi \cdot \phi_i$ measures matching efficiency (i.e., the level of fundamental frictions) in sector i, with Φ denoting the aggregate component and ϕ_i the idiosyncratic sectoral-level component. The number of vacancies and matching efficiency are the only two sources of heterogeneity across sectors in our baseline model.

All existing matches produce Z units of output in every sector. Matches are destroyed exogenously at rate Δ, also common across sectors. Aggregate shocks Z, Δ and Φ, and the vector of vacancies $\mathbf{v} = \{v_i\}$ are drawn from conditional distribution functions $\Gamma_{Z,\Delta,\Phi}\,(Z', \Delta', \Phi'; Z, \Delta, \Phi)$ and $\Gamma_{\mathbf{v}}\,(\mathbf{v}'; \mathbf{v}, Z', \Delta', \Phi')$. The notation shows that we allow for autocorrelation in $\{Z, \Delta, \Phi, \mathbf{v}\}$, and for correlation between vacancies and all the aggregate shocks. The sector-specific matching efficiencies ϕ_i are independent across sectors and are drawn from $\Gamma_\phi\,(\phi'; \phi)$, where $\phi = \{\phi_i\}$. The vector $\{Z, \Delta, \Phi, \mathbf{v}, \phi\}$ takes strictly positive values.

Within each period, events unfold as follows. At the beginning of the period, the aggregate shocks (Z, Δ, Φ), vacancies $\mathbf{v}$, and matching efficiencies ϕ are observed. At this stage, the distribution of active matches $\mathbf{e} = \{e_1, \ldots, e_I\}$ across markets (and hence the total number of unemployed workers u) is also given. Next, unemployed workers choose a labor market i without any impediment to labor mobility. Once the unemployed workers are allocated, the matching process takes place and $h_i = \Phi\phi_i m\,(u_i, v_i)$ new hires are made in each market. Production occurs in the e_i (pre-existing) plus h_i (new) matches. Finally, a fraction Δ of matches are destroyed exogenously in each market i, determining next period's employment distribution $\{e_i'\}$ and stock of unemployed workers u'.

Planner's solution In Appendix A.1 we prove that the planner's optimal rule for the allocation of unemployed workers across sectors in this economy can be written as

$$\phi_1 m_{u_1}\left(\frac{v_1}{u_1^*}\right) = \ldots = \phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) = \ldots = \phi_I m_{u_I}\left(\frac{v_I}{u_I^*}\right), \tag{1}$$

where we have used the "*" to denote the planner's allocation. This condition states that the planner allocates more job seekers to labor markets with more vacancies and higher matching efficiency.

2.2 Heterogeneous productivities and job destructions

We now allow for sector-specific shocks that are uncorrelated across sectors and independent of the aggregate shock.[11] In this extension, we also allow the planner to choose the size of the labor force, but we keep worker separations exogenous. Finally, we allow the planner to choose whether to endogenously dissolve some existing matches and show that, under some conditions, it never chooses to do so. All the derivations for these extensions are contained in Appendix A.2-A.3.

Let labor productivity in sector i be given by $Z \cdot z_i$, where each component z_i is strictly positive, i.i.d. across sectors and independent of Z. Similarly, denote the idiosyncratic component of the exogenous destruction rate in sector i as δ_i. Then, the survival probability of a match is $(1-\Delta)(1-\delta_i)$. It is convenient to proceed under the assumption that $\{Z, 1-\Delta, z_i, 1-\delta_i\}$ are all positive martingales, which amounts to simple restrictions on the conditional distributions $\Gamma_{Z,\Delta,\Phi}, \Gamma_z$, and Γ_δ.[12] Appendix A.2 proves that the planner's optimal allocation rule of unemployed workers equates

$$\frac{z_i}{1-\beta(1-\Delta)(1-\delta_i)}\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) \tag{2}$$

across markets. This rule establishes that the higher vacancies, matching efficiency, and expected discounted productive efficiency in market i, the more unemployed workers the planner wants searching in that market. In particular, expected output of an unemployed worker in sector i is discounted differently by the planner in different sectors because of the heterogeneity in the expected duration of matches.

2.2.1 Endogenous separations

Consider the environment of Section 2.2 and allow the planner to move workers employed in sector i into unemployment or out of the labor force at the end of the period, before choosing the size of the labor force for next period.

[11] This model of sectoral shocks is in the spirit of Lilien (1982). In Section 2.2.2 of Şahin et al. (2012), we laid out an alternative model of sectoral cycles where sectoral productivity fluctuations are driven by the aggregate shock because different sectors have different elasticities to this common factor (in the spirit of Abraham and Katz, 1986). Since results are very similar to the case studied here, we omit the presentation of this variant of the model.

[12] We can allow the vector $x = \{Z, 1-\Delta, z_i, 1-\delta_i\}$ to have the more general linear conditional mean function of the type $\mathbb{E}[x'] = \bar{x} + \rho_x x$. However, the derivations are more convoluted, and we do not make use of this more general assumption in the empirical analysis.

In Appendix A.3 we demonstrate that, if the planner always has enough individuals to pull into (out of) unemployment from (into) out of the labor force, it will never choose to separate workers who are already matched and producing. The planner's allocation rule remains exactly as in equation (2) and all separations are due to exogenous match destructions.

3 Mismatch index and counterfactual unemployment

We now use the planner's allocation rule to derive an index measuring the severity of labor market mismatch between unemployed workers and vacancies. From this point onward we must state an additional assumption, which is well supported by the data as we show below: the individual-market matching function $m\left(u_{it}, v_{it}\right)$ is Cobb-Douglas, i.e.,

$$h_{it} = \Phi_t \phi_{it} v_{it}^{\alpha} u_{it}^{1-\alpha}, \tag{3}$$

where h_{it} are hires in sector i at date t, and $\alpha \in (0,1)$ is the vacancy share common across all sectors.[13] Next, we describe how to use these indexes to construct counterfactuals to measure how much of the recent rise in U.S. unemployment is due to mismatch.

3.1 Mismatch index

Our mismatch index measures the fraction of hires lost because of misallocation, or $(1 - h_t/h_t^*)$ where h_t denotes the observed aggregate hires and h_t^* the planner's hires.

Consider first the benchmark environment of Section 2.1. From (3), summing across markets, the aggregate number of new hires can be expressed as:

$$h_t = \Phi_t v_t^{\alpha} u_t^{1-\alpha} \cdot \left[\sum_{i=1}^{I} \phi_{it} \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} \right]. \tag{4}$$

The optimality condition dictating how to allocate unemployed workers between market i and market j is:

$$\frac{v_{it}}{u_{it}^*} = \left(\frac{\phi_{jt}}{\phi_{it}} \right)^{\frac{1}{\alpha}} \cdot \frac{v_{jt}}{u_{jt}^*}. \tag{5}$$

[13] At this point we abandon the recursive formulation and introduce time t explicitly.

The optimal number of hires that can be obtained by the planner allocating the u_t available unemployed workers across sectors is

$$h_t^* = \Phi_t v_t^\alpha u_t^{1-\alpha} \left[\sum_{i=1}^{I} \phi_{it} \left(\frac{v_{it}}{v_t} \right)^\alpha \left(\frac{u_{it}^*}{u_t} \right)^{1-\alpha} \right]. \tag{6}$$

Substituting the optimality condition (5) in equation (6) , the optimal number of new hires becomes $h_t^* = \Phi_t \bar{\phi}_t v_t^\alpha u_t^{1-\alpha}$, where $\bar{\phi}_t = \left[\sum_{i=1}^{I} \phi_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t} \right) \right]^\alpha$, a CES aggregator of the sector-level matching efficiencies weighted by their vacancy share. Therefore, we obtain an expression for the mismatch index

$$\mathcal{M}_{\phi t} = 1 - \frac{h_t}{h_t^*} = 1 - \sum_{i=1}^{I} \left(\frac{\phi_{it}}{\bar{\phi}_t} \right) \left(\frac{v_{it}}{v_t} \right)^\alpha \left(\frac{u_{it}}{u_t} \right)^{1-\alpha}. \tag{7}$$

$\mathcal{M}_{\phi t}$ measures the fraction of hires lost in period t because of misallocation. This index answers the question: if the planner had u_t available unemployed workers and used its optimal allocation rule, how many additional jobs would it be able to create? These additional hires are generated because, by better allocating the *same number* of unemployed, the planner can increase the aggregate job-finding rate and achieve more hires compared to the equilibrium, which we will call the "direct effect" of mismatch. It is useful to note that, in addition to this direct effect, u_t^* is in general lower than u_t which, for any given allocation rule, translates into a higher aggregate job-finding rate and more hires, which we will call the "feedback" effect. $\mathcal{M}_{\phi t}$ measures only the direct effect of mismatch on hires, but the counterfactual of Section 3.2 fully incorporates the feedback effect as well.

From (7) and (4) one can rewrite the aggregate matching function as

$$h_t = (1 - \mathcal{M}_{\phi t}) \cdot \bar{\phi}_t \cdot \Phi_t v_t^\alpha u_t^{1-\alpha} \tag{8}$$

which makes it clear that higher mismatch lowers the (measured) aggregate efficiency of the matching technology and reduces the aggregate job-finding rate because some unemployed workers search in the wrong sectors (those with relatively few vacancies). The term $\bar{\phi}_t$ can also contribute to a reduction in aggregate matching efficiency when the vacancy shares of the sectors with high ϕ fall.

In Appendix A.4, we show three useful properties of the index. First, $\mathcal{M}_{\phi t}$ is between zero (no mismatch) and one (maximal mismatch). Second, the index is invariant to "pure" aggregate shocks that shift the total number of vacancies and unemployed up or down, but leave the vacancy and unemployment shares across markets unchanged. Third, $\mathcal{M}_{\phi t}$ is increasing in the level of disaggregation. This last property suggests that every statement about the role of mismatch should be qualified with respect to the degree of sectoral disaggregation used.

Consider now the economy of Section 2.2, where labor markets also differ in their level of productive efficiency. It is useful to define "overall market efficiency" as $x_{it} \equiv \phi_{it} z_{it} / \left[1 - \beta\left(1 - \Delta_t\right)\left(1 - \delta_{it}\right)\right]$. Following the same steps, we arrive at the index

$$\mathcal{M}_{xt} = 1 - \sum_{i=1}^{I} \left(\frac{\phi_{it}}{\bar{\phi}_{xt}}\right)\left(\frac{v_{it}}{v_t}\right)^{\alpha}\left(\frac{u_{it}}{u_t}\right)^{1-\alpha}, \tag{9}$$

where

$$\bar{\phi}_{xt} = \sum_{i=1}^{I} \phi_{it}\left(\frac{x_{it}}{\bar{x}_t}\right)^{\frac{1-\alpha}{\alpha}}\left(\frac{v_{it}}{v_t}\right), \text{ with } \bar{x}_t = \left[\sum_{i=1}^{I} x_{it}^{\frac{1}{\alpha}}\left(\frac{v_{it}}{v_t}\right)\right]^{\alpha}. \tag{10}$$

$\bar{\phi}_{xt}$ is an aggregator of the market-level overall efficiencies weighted by their vacancy share.[14]

In the absence of heterogeneity with respect to matching efficiency, productivity, or job destruction, the index becomes $\mathcal{M}_t = 1 - \sum_{i=1}^{I}\left(\frac{v_{it}}{v_t}\right)^{\alpha}\left(\frac{u_{it}}{u_t}\right)^{1-\alpha}$. In what follows, we will also use the notation $(\mathcal{M}_{zt}, \mathcal{M}_{\delta t})$ to denote mismatch indexes for an economy where the only source of heterogeneity is productivity and job destruction rates, respectively.

3.2 Mismatch unemployment

This mismatch index allows us to construct the counterfactual unemployment rate, u_t^*, in the absence of mismatch. The actual aggregate job-finding rate in the economy

[14] The $\mathcal{M}_{xt}$ index still measures the fraction of hires lost because of misallocation. However, since the planner now maximizes output (not employment), theoretically this index could be negative. A version of the index which measures the fraction of *output* (instead of hires) lost to misallocation can be easily computed, and it is always positive.

at date t can be written as

$$f_t = \frac{h_t}{u_t} = (1 - \mathcal{M}_{xt})\, \bar{\phi}_{xt} \Phi_t \left(\frac{v_t}{u_t}\right)^{\alpha}.$$

Let u_t^* be counterfactual unemployment under the planner's allocation rule. The optimal number of hires at date t when u_t^* unemployed workers are available to be allocated across sectors is $\bar{\phi}_{xt} \Phi_t v_t^{\alpha} (u_t^*)^{1-\alpha}$. Therefore, the optimal job-finding rate (in absence of mismatch) is

$$f_t^* = \bar{\phi}_{xt} \Phi_t \left(\frac{v_t}{u_t^*}\right)^{\alpha} = f_t \cdot \underbrace{\frac{1}{(1 - \mathcal{M}_{xt})}}_{\text{Direct Effect}} \cdot \underbrace{\left(\frac{u_t}{u_t^*}\right)^{\alpha}}_{\text{Feedback}} \tag{11}$$

There are two sources of discrepancy between counterfactual and actual job-finding rate. The first term in (11) captures the fact that a planner with u_t available job-seekers to move across sectors would achieve a better allocation and a higher job-finding rate. This effect, which we call the "direct" misallocation effect, is summarized by the mismatch index, as explained. The second term captures a "feedback" effect of misallocation: no mismatch means lower unemployment $(u_t^* < u_t)$ which, in turn, increases the probability of meeting a vacancy for job-seekers. This feedback effect explains why, even if after a period of higher than average mismatch $\mathcal{M}_{xt}$ returns to its average, mismatch unemployment can remain above average for some time, as it takes time for the additional unemployed to be reabsorbed—a pattern we see in our empirical analysis.

Given an initial value for u_0^*, the dynamics of the counterfactual unemployment rate can be obtained by iterating forward on equation

$$u_{t+1}^* = s_t + (1 - s_t - f_t^*)\, u_t^*, \tag{12}$$

where s_t is the separation rate. Our strategy takes the sequences for separation rates $\{s_t\}$ and vacancies $\{v_t\}$ directly from the data when constructing the counterfactual sequence of $\{u_t^*\}$ from (12) , an approach consistent with the theoretical model where vacancy creation and separations are exogenous to the planner.[15]

[15]We avoid the term "constrained efficient" unemployment, because in the extended models of Section 2.1 the planner also controls labor force participation decisions. Therefore, we prefer to interpret

The gap between actual unemployment u_t and counterfactual unemployment u_t^* is mismatch unemployment. This calculation addresses the key question of interest: what is the contribution of mismatch unemployment to the recent rise in the aggregate U.S. unemployment rate? In the rest of the paper we address this question directly.

4 Data

We focus on three major definitions of labor markets: the first is a broad industry classification; the second is an occupation classification, based on both the 2-digit and 3-digit standard occupational classification (SOC) system; the third is a geographic classification, based on U.S. counties and metropolitan areas (MSA's).[16] As discussed in Section 3, our analysis requires information on vacancies, hires, unemployment, productivity, and job separation rates across different labor markets. It also requires market-specific matching efficiency parameters and vacancy share whose calculation involves estimating matching functions.

At the industry level, we use vacancy data from the Job Openings and Labor Turnover Survey (JOLTS), which provides survey-based measures of job openings and hires at a monthly frequency, starting from December 2000, for seventeen industry classifications.[17] At the occupation and county level, we use vacancy data from the Help Wanted OnLine (HWOL) dataset provided by The Conference Board (TCB). This is a novel data series that covers the universe of online advertised vacancies posted on internet job boards or in newspaper online editions. It covers roughly 16,000 online job boards and provides detailed information about the characteristics of advertised vacancies for between three and four million unique active ads each month.[18] The HWOL database started in May 2005 as a replacement for the Help-Wanted Advertising Index of print advertising maintained by TCB.[19]

u_t^* as the counterfactual unemployment rate under the planner's allocation rule of unemployed workers across sectors, abstracting from possible discrepancies between the planner's labor force participation choice and the corresponding equilibrium outcome observed in the data.

[16] See Tables B1-B3 in Appendix B for a list of industry and occupation classifications used in the empirical analysis.

[17] Since the JOLTS is a well known and widely used survey, we do not provide further details. For more information, see http://www.bls.gov/jlt/. See also Faberman (2009).

[18] The data are collected for The Conference Board by Wanted Technologies. For detailed information on survey coverage, concepts, definitions, and methodology see the Technical Notes at http://www.conference-board.org/data/helpwantedonline.cfm

[19] Our empirical analysis covers the December 2000-June 2011 period for the JOLTS, and May

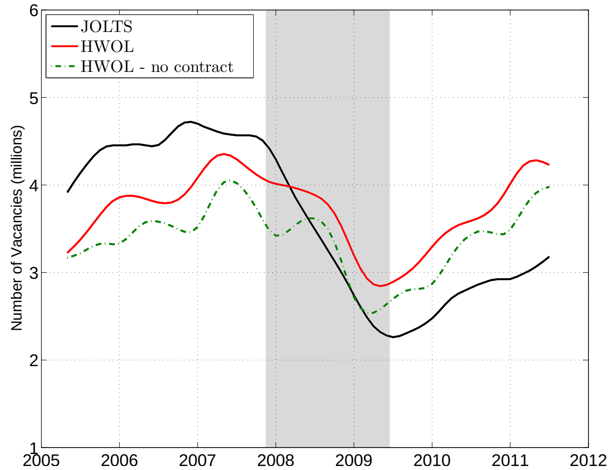


Figure 1: Comparison between the JOLTS and the HWOL (The Conference Board Help Wanted OnLine Data Series) aggregate time series.

Each observation in the HWOL database refers to a unique ad and contains information about the listed occupation at the 6-digit level, the geographic location of the advertised vacancy down to the county level, whether the position is for full-time, part-time, or contract work (essentially self-employed contractors or consultants: e.g., computer specialists, accountants, auditors), the education level required for the position, and the hourly and annual mean wage.[20] For 57% of ads we also observe the industry NAICS classification. The majority of online advertised vacancies are posted on a small number of job boards: about 60% of all ads appear on five job boards.[21]

It is worth mentioning some measurement conventions in the HWOL data: first, the same ad can appear on multiple job boards. To avoid double-counting, TCB uses a sophisticated unduplication algorithm that identifies unique advertised vacancies on the basis of the combination of company name, job title/description, city or state. Second, there are some cases in which multiple locations (counties within a state) are

2005-June 2011 for the HWOL.

[20]The education and wage information is imputed by TCB. Education is imputed from BLS data on the education content of detailed 6-digit level occupations. Wages are imputed using BLS data from the Occupational Employment Statistics (OES), based on the occupation classification. For a subset of the ads we also observe the sales volume, the number of employees of the company, and the actual advertised salary range, but in this paper we do not attempt to use this additional information.

[21]The five largest job boards are: CareerBuilder, Craigslist, JOBcentral, Monster, and Yahoo!HotJobs.

listed in a given ad for a given position. TCB follows the rule that if the counties are in the same state or MSA the position is taken to represent a single vacancy, but if they appear in different MSA's and in different states they reflect distinct vacancies. In addition, the dataset records one vacancy per ad. There is a small number of cases in which multiple positions are listed, but the convention used is one vacancy per ad.

More importantly, the growing use of online job boards over time may induce a spurious upward trend. Figure 1 plots JOLTS vacancies and HWOL ads at the national level. The total count of active vacancies in HWOL is below that in JOLTS until the beginning of 2008 and is above from 2008 onwards, a pattern which may reflect the increasing penetration of online job listings over time. Nevertheless, the average difference between the two aggregate series is only about 16% of the JOLTS total, and the correlation between the two aggregate series is about 0.65. To the extent that this trend towards online recruitment does not differ too much across sectors, our calculations are not affected. In Section 6, we propose a reweighing scheme for HWOL that aligns it more closely to JOLTS and show that our findings remain robust. We report additional detailed comparisons between the JOLTS and HWOL vacancy series in Appendix B.1.

We calculate unemployment counts from the Current Population Survey (CPS) for the same industry and occupation classification that we use for vacancies.[22] For geography, we use the Local Area Unemployment Statistics (LAUS) which provides monthly estimates of total unemployment at the county and MSA level.[23] The CPS reports the industry and occupation of unemployed workers' previous jobs. We begin by assuming that all unemployed workers search only in the sector that they had last worked in. We relax this assumption in Section 6. The small sample size of the CPS limits the level of disaggregation of our analysis, and prevents us from using HWOL ads data to their full effect.[24]

We use various proxies for productivity, depending on data availability. At the industry level, we compute labor productivity by dividing value added for each in-

[22]Industry affiliations are not available for all unemployed workers in the CPS. From 2000-2010, on average about 13.3% of unemployed do not have industry information. Only about 1.5% of unemployed are missing occupation information. Some of these workers have never worked before and some are self-employed.

[23]See http://www.bls.gov/lau/ for more information on LAUS.

[24]The average number of unemployed in the CPS for the May 2005 to June 2011 period is 4,557 with a range of 2,808 to 12,436.

dustry from the Bureau of Economic Analysis (annual data) by average employment in that industry from the Establishment Survey.[25] At the occupation level, for lack of a better proxy, we use annual data on average hourly wages from the Occupational Employment Statistics (OES).[26] Similarly, at the county level, we use median weekly wage earnings from the Quarterly Census of Employment and Wages (QCEW).[27] We recognize that wage levels might be affected by factors other than productivity like unionization rates, compensating differentials, monopoly rents, etc. To partially address this issue, we normalize the average wage for each occupation to unity at the beginning of our sample and focus on relative wage movements over time. We also apply the same normalization to industry-level productivity measures for consistency.

We calculate job destruction rates at the industry level from the Business Employment Dynamics (BED) as the ratio of gross job losses to employment.[28] Since the BED is quarterly, we assume that the destruction rate is the same for the three months corresponding to a specific quarter and impute the corresponding monthly destruction rates. Because job destruction rates by occupation are not available, we compute the employment to unemployment transition rates by occupation in the last job from the CPS semi-panel. Figures B3 and B4 in the appendix show the evolution of productivity and job destruction rates for selected industries and occupations.

To compute market-specific matching efficiency parameters, ϕ_i, and vacancy share α, we estimate aggregate and sector-specific (constant-returns to scale) matching functions using various specifications, estimation methods, and data sources. Appendix B.2 contains a detailed description of methodology and results. Our findings indicate that a value of the vacancy share $\alpha = 0.5$ is appropriate.[29] Tables B6-B8 in Appendix B contain estimates of sector-specific matching efficiencies.

[25]http://www.bea.gov/industry/

[26]See http://www.bls.gov/oes/

[27]See http://www.bls.gov/cew/

[28]See http://www.bls.gov/bdm/. We recognize this is an imperfect proxy for separations, but (i) monthly employment-unemployment transitions computed from CPS semi-panel at the industry level are much noisier, and (ii) during 2001-2010, only 16 pct of quits ends into unemployment, as opposed to 91 pct of layoffs (see Elsby et al., 2010).

[29]This value is roughly in the middle of the range of estimates used in other recent papers in the matching literature. A few examples are $\alpha = 0.5$ in Davis, Faberman, and Haltiwanger (2010), $\alpha = 0.28$ in Shimer (2005), $\alpha = 0.54$ in Mortensen and Nagypal (2007), α between 0.66 and 0.72 in Barnichon and Figura (2011). Moreover, our mismatch indices are typically highest for $\alpha = 0.5$; therefore, this value is consistent with the spirit of reporting an upper bound for mismatch unemployment.

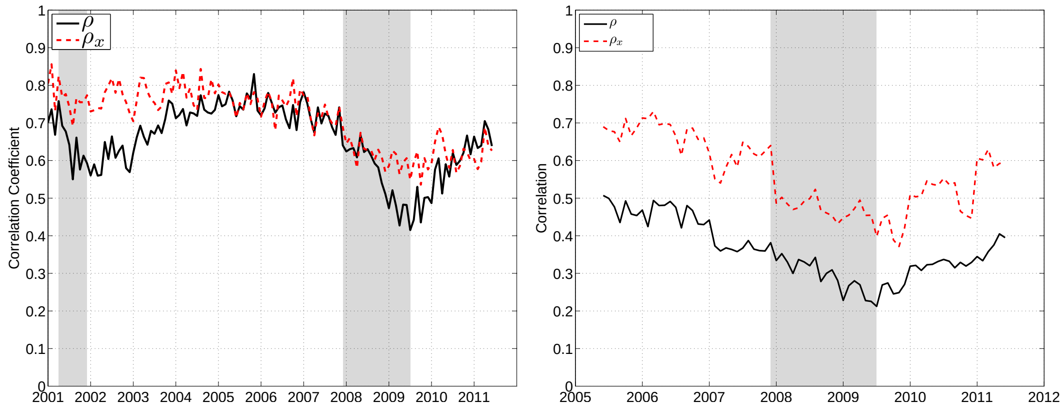


Figure 2: Correlation coefficient between u and v shares across industries (left panel) and two digit occupations (right panel).

5 Results

5.1 Industry-level mismatch

From our definition of mismatch, it is clear that there is a close association between mismatch indexes and the correlation between unemployment and vacancy shares across sectors. The planner's allocation rule implies a perfect correlation between unemployment shares and (appropriately weighted) vacancy shares. A correlation coefficient below one is a signal of mismatch, and a declining correlation is a signal of worsening mismatch. Figure 2 (left panel) plots the time series of this correlation coefficient across industries over the sample period. In particular, we report two different correlation coefficients motivated by the definitions of the mismatch indexes we derived in Section 3: 1. ρ: between (u_{it}/u_t) and (v_{it}/v_t) and 2. ρ_x: between (u_{it}/u_t) and $(x_i/\bar{x}_t)^{\frac{1}{\alpha}}(v_{it}/v_t)$. The two series behave very similarly. They drop sharply from early 2006 to mid 2009 and recover thereafter, indicating a rise in mismatch during the recession that is, however, relatively short-lived.[30]

[30]It is also useful to examine the evolution of vacancy and unemployment shares of different industries. In Figure B5, we plot the vacancy and unemployment shares for a selected set of industries using the JOLTS definition in the Appendix. The shares have been relatively flat in the 2004-2007 period. However, starting in 2007, vacancy shares started to change noticeably. Construction and durable goods manufacturing were among the sectors which experienced a decline in their vacancy shares while the health sector saw its vacancy share increase. Concurrently, unemployment shares of construction and durable goods manufacturing went up while the unemployment share of the health sector decreased. Starting from 2010, sectoral unemployment and vacancy shares began to regress towards their pre-recession levels, with the exception of the construction sector. The vacancy share of

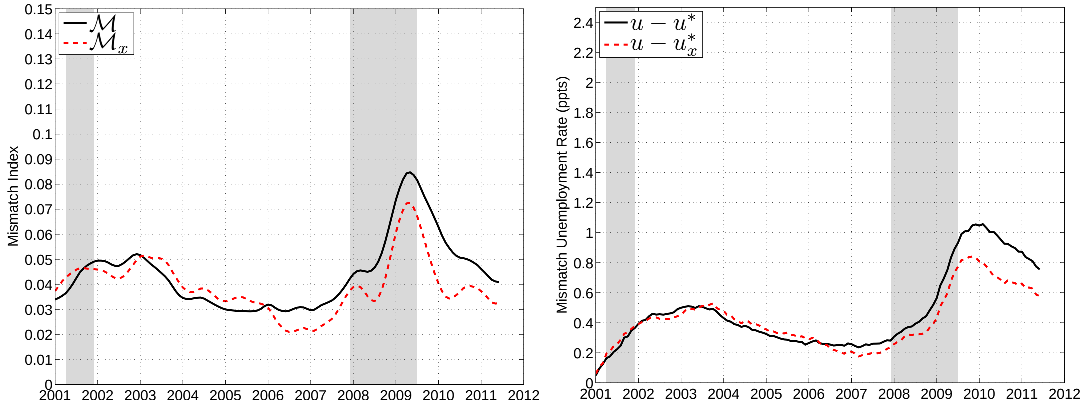


Figure 3: Mismatch index $\mathcal{M}_t$ and $\mathcal{M}_{xt}$ by industry (left panel) and the corresponding mismatch unemployment rates (right panel).

The left panel of Figure 3 plots the unadjusted index, $\mathcal{M}_t$ and the one adjusted for heterogeneity, $\mathcal{M}_{xt}$.[31] This figure shows that, before the last recession (in mid 2006), the fraction of hires lost because of misallocation of unemployed workers across industries ranged from 2-3 percent per month, depending on the index used. At the end of the recession, in mid 2009, it had increased to roughly 7-8 percent per month, and it has since dropped again to almost its pre-recession level. To sum up, both indexes indicate a sharp rise in mismatch between unemployed workers and vacant jobs across industries during the recession, and a subsequent fairly rapid decline.

How much of the observed rise in the unemployment rate can be explained by mismatch? The right panel of Figure 3 shows mismatch unemployment (i.e., the difference between the actual and the counterfactual unemployment rates) at the industry level for the 2001-2011 period, computed as described in Section 3.2. Table 1 shows the change in mismatch unemployment between October 2009 and the average of 2006.[32] The main finding is that worsening mismatch across these seventeen industries explains between 0.59 and 0.75 percentage points of the rise in U.S. un-

the construction sector remains well below its pre-recession level.

[31]Note that all mismatch indexes throughout the paper are HP filtered to eliminate high frequency movements and better visualize the variation in the indexes. To facilitate the comparison across different definitions of labor markets, we plot all the mismatch indexes and mismatch unemployment rates using the same vertical distance on the y axis, 0.15 and 2.5 percentage points, respectively.

[32]Note that the average unemployment rate was 4.6% in 2006 and 10.0% at its peak in October 2009, indicating a 5.4 percentage point increase. Throughout the paper we compare the average of 2006 with the unemployment peak (October 2009) when we discuss the role of mismatch in the increase in the unemployment rate.

	Index	$u_{06}-u^*_{06}$	$u_{10.09}-u^*_{10.09}$	$\Delta(u-u^*)$	$\Delta(u-u^*)/\Delta u$
Industry	$\mathcal{M}$	0.26	1.01	0.75	13.9%
	$\mathcal{M}_x$	0.24	0.84	0.59	11.0%
	$\mathcal{M}_{u-adj}$	0.25	0.89	0.65	11.9%
	$\mathcal{M}_x^{v^*}(\varepsilon=0.5)$	0.67	1.90	1.21	22.5%
	$\mathcal{M}_x^{v^*}(\varepsilon=1.0)$	0.34	1.24	0.90	16.6%
	$\mathcal{M}_x^{v^*}(\varepsilon=2.0)$	0.26	0.95	0.69	12.7%
2-digit Occupation	$\mathcal{M}$	0.85	2.00	1.14	21.3%
	$\mathcal{M}_x$	0.42	1.02	0.60	11.1%
	$\mathcal{M}_{u-adj}$	0.84	2.00	1.16	21.4%
	$\mathcal{M}_{v-adj}$	0.93	2.12	1.19	22.1%
	$\mathcal{M}_x^{v^*}(\varepsilon=0.5)$	1.08	2.60	1.52	28.1%
	$\mathcal{M}_x^{v^*}(\varepsilon=1.0)$	0.74	1.81	1.07	19.7%
	$\mathcal{M}_x^{v^*}(\varepsilon=2.0)$	0.58	1.41	0.83	15.3%
3-digit Occupation	$\mathcal{M}$	1.33	2.91	1.58	29.3%
	$\mathcal{M}_x$	0.79	1.73	0.94	17.4%
Routine/Cognitive	$\mathcal{M}_{RC}$	0.41	1.07	0.67	12.3%
County	$\mathcal{M}$	0.32	0.46	0.14	2.6%
	$\mathcal{M}_z$	0.32	0.45	0.13	2.5%
2-digit $\times$ division (quarterly)	$\mathcal{M}$	0.81	1.71	0.90	16.9%
2-digit (quarterly)	$\mathcal{M}$	0.68	1.53	0.85	16.0%

Table 1: Changes in mismatch unemployment at the industry, occupation, and county levels. All the differences are calculated as the difference between October 2009 and the average of 2006. Note that $\Delta u = 5.4$ percentage points.

employment from 2006 to its peak, i.e., at most 14 percent of the increase. Mismatch unemployment has declined since early 2010, but remains above its pre-recession levels.

Appendix B.3 contains a sensitivity analysis on industry-level mismatch with respect to (i) values of α ranging from 0.3 to 0.7; (ii) alternative estimates of matching efficiency ϕ_i's; and (iii) HWOL vacancy data by industry. Results are very robust: the contribution of mismatch to the rise in the unemployment rate around the Great Recession varies between 0.52 and 0.88 percentage points.

5.2 Occupation-level mismatch

We now present our results on mismatch between vacancies and unemployment across 2- and 3-digit occupations. Recall that the HWOL ads data used for these

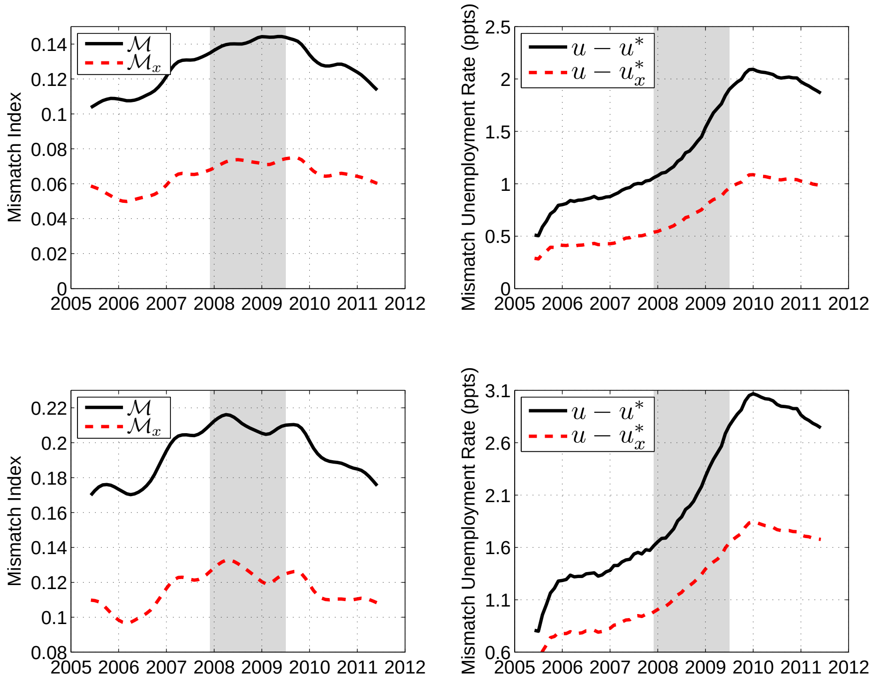


Figure 4: Mismatch indexes $\mathcal{M}_t$ and $\mathcal{M}_{xt}$ by 2-digit occupation (upper left panel) and 3-digit occupation (lower left panel). Corresponding mismatch unemployment rates for 2-digit (upper right panel) and 3-digit occupations (lower right panel).

calculations begin in May 2005 and the latest observation is June 2011.

Figure 2 (right panel) plots the correlation between vacancy and unemployment shares across 2-digit SOC's. As for the industry-level analysis, we document a significant decline for both measures during the recession and a subsequent pick-up starting in mid-2009.[33]

Figure 4 plots the $\mathcal{M}_t$ and $\mathcal{M}_{xt}$ indexes (left panels) and the resulting mismatch unemployment (right panels) for 2 and 3-digit SOC's.[34] The raw $\mathcal{M}_t$ index for 2-

[33]Figure B10 in the Appendix shows the unemployment and vacancy shares of selected 2-digit SOC's. As the figure indicates, the shares have changed noticeably during the most recent downturn. Business and financial operations, production and construction/extraction were among the occupations that experienced a decline in their vacancy shares and an increase in their unemployment shares. Concurrently, vacancy shares of health-care practitioner and sales and related occupations went up and the corresponding unemployment shares declined. Starting from 2010, similar to the JOLTS data, unemployment and vacancy shares began to normalize.

[34]There are 22 2-digit SOC's and 93 3-digit SOC's. We use all the 2-digit categories with the exception of Farming, Fishing, and Forestry. We exclude 3-digit SOC's that exhibit fewer than 10 observations in the CPS unemployment counts at least once in the sample period. These small cells account for 60% of the 3-digit SOC's, but represent only 15.6% of unemployed workers in the CPS.

digit occupations rises by almost 4 percentage points. Similar to the pattern observed for industries, the rise in mismatch leads the recession by over a year. As seen in the figure and in Table 1, based on the $\mathcal{M}_t$ index, around 1.1 percentage points (or around 21%) of the recent surge in U.S. unemployment can be attributed to occupational mismatch measured at the 2-digit occupation level. At the 3-digit level, the portion of the increase in unemployment attributable to mismatch is around 1.6 percentage points (or roughly 29% of the rise in the unemployment rate).

The efficiency-weighted $\mathcal{M}_{xt}$ index is lower than the unadjusted index and features a smaller rise, implying around 2% of additional hires lost because of mismatch. This index suggests that between 0.6 and 0.9 percentage points of the rise in the unemployment rate (or between 11% and 17% of the increase) was due to mismatch at the 2-digit and 3-digit SOC levels, respectively. Therefore, similar to what we found for industries, the index weighted by matching and productive efficiency implies a smaller role for mismatch unemployment.

Appendix B.3 contains a sensitivity analysis on occupational-level mismatch with respect to (i) values of α; (ii) alternative estimates of matching efficiency ϕ_i's. The findings reported here are very robust.

5.2.1 The role of job polarization for occupational mismatch

Job polarization refers to the increasing concentration of employment in the highest- and lowest-wage occupations, with job opportunities in middle-skill occupations disappearing, as documented by Acemoglu and Autor (2011). To capture the effect of job polarization on mismatch, we classify 2-digit occupations into four categories: routine cognitive, routine manual, non-routine cognitive, and non-routine manual. We call this classification "Routine/Cognitive" and denote the index with $\mathcal{M}_{RC}$.[35] Figure B13 in Appendix B shows the unadjusted mismatch index across these four occupation groups as well as the index calculated at the 2-digit level, and the corresponding mismatch unemployment rates. Our findings are summarized in Table 1. The lower level of the index suggests additional mismatch within these four broad categories. Despite the gap in the level of the two indices, the behavior of the $\mathcal{M}_{RC}$

[35] We classify occupations at the 2-digit level instead of directly using Acemoglu and Autor's classification. While their way of classifying occupations is more detailed, our classification broadly captures this distinction and is more comparable with the rest of our analysis. See Table B2 in the Appendix for our classification of occupations into these four groups.

index is very close to the occupational mismatch index $\mathcal{M}_{\sqcup}$ computed using 2-digit occupations. In essence, the vacancy (unemployment) share dropped (rose) faster for routine manual occupations relative to the other groups, accounting for at least half of the increase in mismatch unemployment across the twenty-one 2-digit occupations.

Jaimovich and Siu (2012) link the job polarization hypothesis to jobless recoveries by analyzing employment changes during recessions and recoveries across these occupational groups. They show that employment declined more in routine occupations during the most recent downturn.[36] This relatively large decline is in line with the increase in mismatch during the recession. They also show that employment remained stagnant in all occupational categories during the recovery, which is consistent with the decline in mismatch after the recession.

5.2.2 Occupational mismatch within education groups and within regions

Is occupational mismatch a more relevant source of unemployment dynamics for less skilled or for more skilled workers? A priori, the answer is ambiguous: more education means more adaptability, but also more specialized knowledge. To address this question, we define four education categories (less than high school diploma, high school diploma or equivalent, some college or Associate's degree, Bachelor's degree or higher) and analyze mismatch by 2-digit occupation *within* each of these four education groups.

As noted before, each job listing recorded in HWOL constitutes an individual observation with a 6-digit occupation classification. The BLS provides information on the distribution of workers employed in each 6-digit occupation broken down by their educational attainment.[37] We allocate the total count of vacancies from HWOL in a given month for a given 6-digit occupation to each of the four education groups we consider, proportionally to the educational attainment distributions from the BLS.[38] Finally, we aggregate up to the 2-digit level to obtain vacancy counts for each occupation by education cell. The CPS provides information on the education level of the

[36] See the bottom panel of Figure 6 in Jaimovich and Siu (2012), p. 12.

[37] This information comes from the American Community Survey microdata from 2006-08. See the BLS website at http://www.bls.gov/emp/ep_table_111.htm; see also http://www.bls.gov/emp/ep_education_tech.htm for additional details.

[38] For robustness, we have also experimented with other allocation rules, for instance not imputing vacancies of a given 6-digit SOC to an education level that accounts for less than 15% of the workers in that occupation. The results are very similar.

	$u_{06} - u^*_{06}$	$u_{10.09} - u^*_{10.09}$	$\Delta(u - u^*)$	Δu	$\Delta(u - u^*)/\Delta u$
Less than High School	0.71	1.69	0.98 ppts	8.5 ppts	11.5%
High School Degree	0.60	1.50	0.89 ppts	6.9 ppts	12.9%
Some College	0.71	1.68	0.97 ppts	5.3 ppts	18.2%
College Degree	0.38	1.03	0.65 ppts	2.7 ppts	23.9%

Table 2: Changes in mismatch unemployment across 2-digit occupations for different education groups. All the changes are calculated as the difference between October 2009 and the average of 2006. Note that $\Delta u = u_{10.09} - u_{06}$ and that Δu varies by education.

unemployed.

The counterfactual exercises summarized in Table 2 reveal a very clear pattern: the contribution of occupational mismatch to the rise in unemployment between 2006 and 2010 grows as we move from the lowest to the highest education category. In particular, for the less than high school group, mismatch explains a little less than one percentage point (12%) of the 8.5 percentage point increase in unemployment for that group. For high school graduates, mismatch explains 0.89 (13%) out of the 6.9 percentage point increase in unemployment. For those with some college, mismatch explains about 1.0 (18%) out of a 5.3 percentage point rise in unemployment, and for college graduates 0.65 (24%) out of the 2.7 percentage point observed increase. Thus, the fraction of the rise in unemployment that can be attributed to the rise in occupational mismatch increases monotonically with education from about one eighth to roughly one quarter of the increase for each group.[39]

Looking at occupational mismatch separately for each of the four U.S. Census regions (Figure 5) reveals that the only region where our index is still significantly above its pre-recession level is the West, i.e. the region where the fall in house prices and the rise in unemployment were the sharpest.

5.3 Geographical mismatch

We perform our geographical analysis on mismatch across U.S. counties using the HWOL data on online job ads coupled with LAUS data on the unemployed. We focus on counties whose population is at least 50,000 and group together counties in the

[39]Figures B14 in the Appendix plots mismatch indexes within each broad education category. The index for college graduates is the only one which is still significantly above its 2006 level.

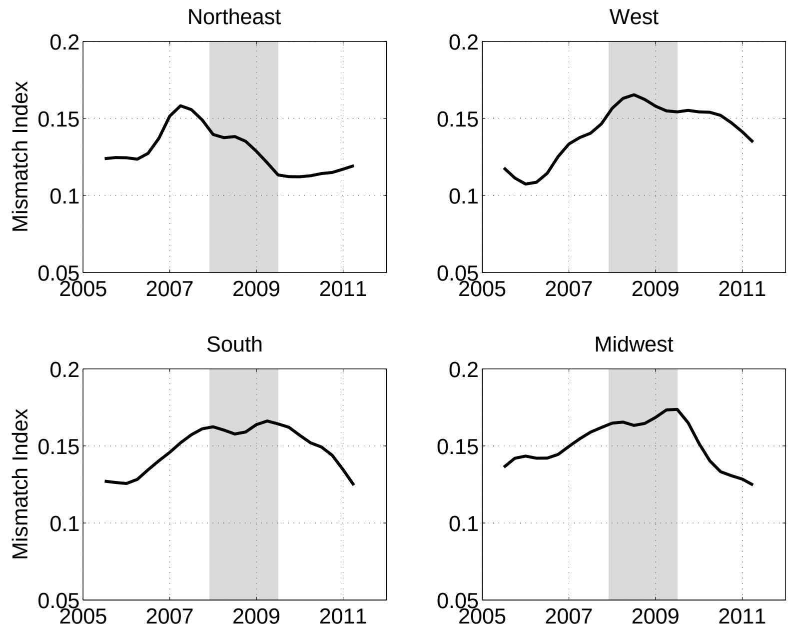


Figure 5: 2-digit occupational mismatch indexes $\mathcal{M}_t$ in the four U.S. Census regions.

same metropolitan area to capture the notion of local labor markets. The procedure gives a total of 280 local labor markets.[40]

Figure 6 shows the mismatch indexes $\mathcal{M}_t$ and $\mathcal{M}_{zt}$ and the corresponding mismatch unemployment rates. We find that geographic mismatch is very low (about one fifth of the size of the index for 3-digit occupations, even though the number of active sectors is much higher) and is essentially flat over the sample period under consideration. In particular, this measure of mismatch does not display a marked cyclical pattern, indicating that counter-cyclicality is not a mechanical feature of our index, but it depends on how the distribution of unemployment and vacancies across sectors evolves over the cycle.

Unsurprisingly, the rise in mismatch unemployment according to this index is around one tenth of a percentage point, implying that geographical mismatch—across U.S. counties and MSAs—played a negligible role in the recent dynamics of U.S. unemployment. This finding is consistent with other recent work that investigated the link between housing market and labor market using different methods (see, e.g.,

[40]We also compute geographic mismatch for the 50 U.S. states using the HWOL data on online job ads coupled with CPS data on the unemployed. The JOLTS provides limited geographic information, enabling us to study mismatch only across the four broad Census regions. Our conclusions from these state- and region-based analyses are fully aligned with the county-based study.

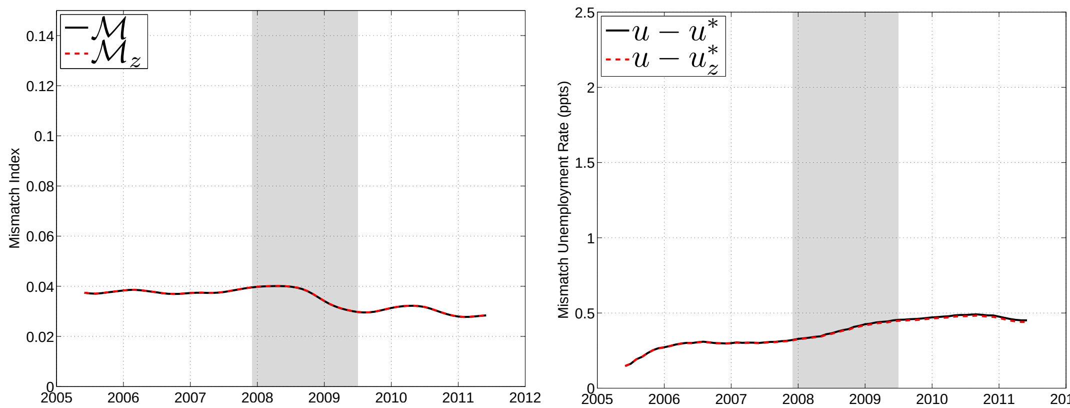


Figure 6: Geographical mismatch indexes $\mathcal{M}_t$ and $\mathcal{M}_{zt}$ by county (left panel) and corresponding mismatch unemployment rates (right panel).

Schulhofer-Wohl, 2010, and Farber, 2012).[41]

We also examine mismatch defining labor markets as a combination of occupations and locations. Because of the small sample size of the CPS, we define labor markets as the interaction of 2-digit occupations and the nine Census divisions. The resulting mismatch indexes and mismatch unemployment are presented in Figure B15 and Table 1. The dynamics of both mismatch index and the mismatch unemployment are very similar to those computed at the 2-digit occupation level.[42]

5.4 Is the Great Recession different from the 2001 recession?

For our industry-level analysis, we are able to compare the evolution of mismatch unemployment in the Great Recession to that of the 2001 recession. Figures 2 and 3 show that the fall in the cross-sectoral unemployment-vacancy correlation and the rise in our mismatch index is common to the last two downturns. In Table B11 we report our calculations on the role of mismatch unemployment in 2001. We find that worsening mismatch accounted for a larger portion of the (smaller) rise in unemployment in the 2001 recession (23% instead of 11-14%). This finding echoes the fact

[41] A related concern regarding geographic mobility is the apparent observation that the rate of interstate migration in the U.S. reached a postwar low. However, Kaplan and Schulhofer-Wohl (2010) show that this is largely a statistical artifact arising from a change in survey procedures for missing values. After removing this spurious effect, they find that the annual interstate migration rate follows a smooth downward trend from 1996 to 2010.

[42] Due to the small sample, we compute this index with quarterly data. To facilitate the comparison, we also compute $\mathcal{M}_t$ for 2-digit occupations at the quarterly frequency.

that the dynamics of employment for different occupational groups were much more asymmetric in 2001 than in 2008, as documented by Jaimovich and Siu (2012).

6 Robustness on job-seeker and vacancy measures

There are three potential sources of bias that might affect our mismatch measures. The first is the assumption that each unemployed worker is searching in the same industry or occupation as the one where she was last employed. Second, our unemployment counts do not include discouraged workers. Since some workers get discouraged from job search and drop out of the labor force temporarily during periods of high unemployment, we might underestimate the true number of potential job seekers in some sectors. Third, HWOL data on aggregate vacancies show a stronger upward trend than their JOLTS counterpart. If this trend is uneven across sector, it may bias our mismatch measures. In this section, we verify the robustness of our findings to these measurement issues.

6.1 Adjustment for direction of search

The number of unemployed workers *searching* for jobs in a particular sector does not necessarily coincide with the number of workers whose last employment was in that sector. Here, we propose an alternative calculation of the number of job-seekers in each industry or occupation by exploiting the semi-panel dimension of the CPS. Since respondents in the CPS are interviewed for several consecutive months, we can track unemployed workers who find new employment from one month to the next and record: 1. industry (and occupation) of the job prior to the worker's unemployment spell; 2. industry (and occupation) of the new job. We then create annual transition matrices (from sector i to sector j) by aggregating monthly flows.[43] We then infer the number of job seekers in each sector using a simple statistical algorithm, whose key assumption is that every unemployed searching for a job in sector j has the same probability of being hired, independently of the sector of origin, except when coming from sector j itself in which case she is allowed to have a higher job-finding rate. The method is outlined in detail in Appendix B.4.

[43] In implementing this procedure, we closely follow Hobijn (2012).

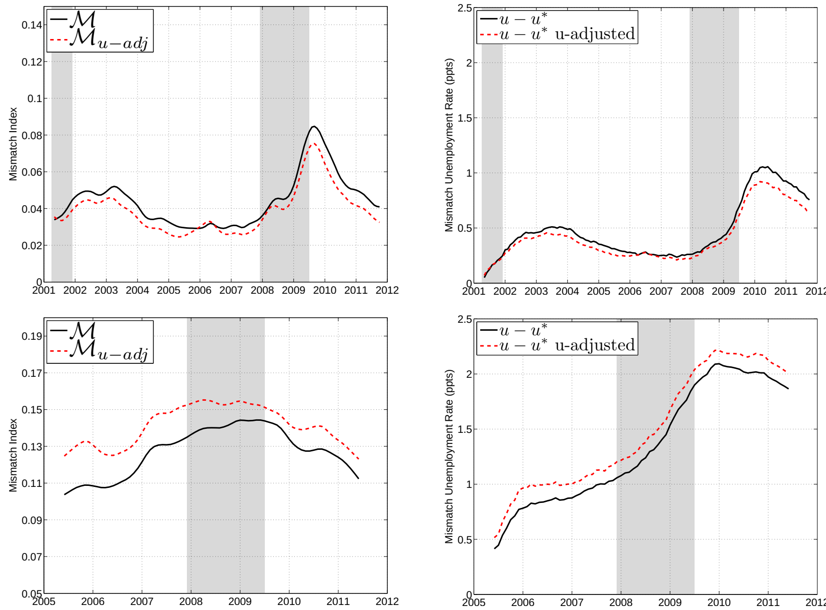


Figure 7: Mismatch index with unadjusted ($\mathcal{M}$) and adjusted ($\mathcal{M}_{u-adj}$) unemployment counts by industry (top-left panel) and corresponding mismatch unemployment rates (top-right panel). Mismatch index with adjusted and unadjusted unemployment counts by occupation (bottom-left panel) and corresponding mismatch unemployment rates (bottom-right panel).

We first report our results by industry. The top-left panel of Figure 7 shows the mismatch index calculated using the adjusted unemployment counts, which we call $\mathcal{M}_{u-adj}$, as well the unadjusted $\mathcal{M}$ index. The adjustment causes the level of the index to decrease somewhat during the sample period. When using the adjusted counts, 0.65 percentage points of the roughly 5.4 percentage point rise in the U.S. unemployment rate is due to industry-level mismatch, compared to 0.75 percentage points without the adjustment (top-right panel).

The bottom row of Figure 7 reports our analysis by occupation. Again, the behavior of the adjusted $\mathcal{M}_{u-adj}$ index and of the resulting mismatch unemployment is very similar to the case without adjustment. In contrast to the industry-level analysis, the adjusted index for occupations is slightly higher than in the baseline case. Quantitatively, the contribution of mismatch to the rise in the U.S. unemployment rate is virtually the same when using adjusted unemployment counts by occupation. Table 1 summarizes these results.

6.2 Adjustment for discouraged workers

According to the CPS, an individual is unemployed if he or she does not have a job, has actively looked for employment in the past four weeks and is currently available to work. However, it is possible that some workers get discouraged from unsuccessful job search and drop out of the labor force temporarily during periods of high unemployment.[44] If workers from certain occupations or industries are more likely than others to get discouraged, our mismatch measures may be biased. For example, if a high fraction of the unemployed whose prior occupation was construction-related drop out of the labor force and stay on the sidelines (to re-enter the labor force at a later stage), the number of unemployed construction workers is an under-estimate of the true number of potential job seekers in the construction sector. In this example, actual mismatch would be larger.

We calculate the unemployment to "discouraged not in the labor force" (UD) flow rates conditional on workers' previous occupations and industries. Tables B12 and B13 in Appendix B show that the rates at which unemployed workers flow into the discouraged worker state are similar across industries and occupations. As a consequence, adjusting the unemployment counts by including discouraged workers affects the unemployment shares of different industries and occupations only marginally. As a robustness check, we recomputed the basic mismatch index $\mathcal{M}_t$ using this extended definition of unemployment where we include workers who flow from unemployment to discouragement. The difference between the modified mismatch index and the original index is quantitatively insignificant (a difference of 0.0002 on average).[45]

[44] The Current Population Survey classifies as discouraged workers those individuals "not in the labor force who want and are available for a job and who have looked for work sometime in the past 12 months (or since the end of their last job if they held one within the past 12 months), but who are not currently looking because they believe there are no jobs available or there are none for which they would qualify."

[45] As a further check, we repeate this calculation by including all the workers who flow from unemployment to non-participation (UN). Tables B12 and B13 report also the UN flows by industry and occupation. Results are very similar.

6.3 Reweighing of HWOL vacancies

The two main concerns with the HWOL data are that (i) some sectors may systematically over- or under-use online recruitment tools compared to the aggregate and (ii) the upward trend in the penetration of online advertisement may be faster or slower in some sectors than others. To address these concerns, we reweight HWOL vacancy counts by occupation in order to match the total vacancy counts by industry and region in JOLTS, month by month. Appendix B.5 describes our approach in detail.

Table B14 reports the estimated weights by industry and region. A low (high) weight means that sector or region makes use of online recruitment boards more (less) than the aggregate economy. Our findings are quite intuitive: Finance, Real Estate, and Professional Services are among the most over-represented industries in online recruitment, and Accommodation, Government, and Construction among the most under-represented. Weights change somewhat over time, but the correlation between the 2005-06 and the 2010-11 weights is 0.90, indicating that the upward trend is quite common across sectors.

Figure B18 shows that, when we recompute the mismatch index using these reweighted vacancy counts by 2-digit occupation ($\mathcal{M}_{v-adj}$) we do find a slightly higher increase in occupational mismatch, but as can be seen in Table 1, the counterfactual exercise yields results similar to our baseline calculation with the raw HWOL data.[46]

7 Endogenous vacancy distribution

In this section, we relax the assumption of exogeneity of the distribution of vacancies maintained so far. Why would endogenizing vacancies affect our calculations? If, in equilibrium, too many job-seekers search in the sectors with low matching

[46]We also redid our analysis using separate weights by region or by industry: our findings are very similar to the baseline results. In a previous version of the paper (Şahin et al, 2012) we also address the issue that vacancies may be measured with error (in both JOLTS and HWOL), since not all hires occur through formal advertisement (see, e.g., Galenianos, 2012, for an analysis of hiring through referrals). We show that markets where vacancies are severely under-reported look like markets with higher matching efficiency, and argue that our calculations are still appropriate. Intuitively, it makes no difference to the planner whether ϕ_i is high in a sector because pure matching efficiency is high or because actual vacancies are larger than those formally advertised: in both cases, the planner would like to allocate many job-seekers to that sector.

and productive efficiency, private firms' job creation decisions are distorted: an excessive number of vacancies will be posted in those sectors (because of the higher probability of recruitment) compared to the choice of a planner who allocates vacancies and job seekers purely based on relative efficiency across sectors. The result is a lower number of aggregate vacancies and a lower aggregate job-finding rate in equilibrium—another "feedback" effect of mismatch stemming, this time, from the vacancy side.

We begin by stating some additional assumptions on the equilibrium data generating process needed to measure the cost of vacancy creation. We then proceed to explain formally this additional feedback effect of mismatch. Finally, we present our findings. Appendix A.5 contains more details on all the derivations.

7.1 Measurement of the vacancy creation cost

Let the cost, in terms of final good, of creating v_i vacancies in sector i be

$$K_i\left(v_i\right) = \kappa_i^{\varepsilon} \cdot \frac{v_i^{1+\varepsilon}}{1+\varepsilon}, \tag{13}$$

with $\varepsilon \in (0, \infty)$ to guarantee convexity of the K_i function.[47] With this isoelastic specification, ε measures the elasticity of vacancy creation, i.e., how the (log of the) the marginal cost increases with the (log of the) number of vacancies. The variable κ_i shifts the cost of vacancy creation across sectors and over time. We let κ_i be i.i.d. across sectors and independent of the other idiosyncratic shocks, and denote its conditional distribution as $\Gamma_\kappa\left(\kappa', \kappa\right)$.[48] The choice of how many vacancies to post takes place before the allocation of unemployment across sectors, but after observing sectoral and aggregate states.

The first challenge we face is how to estimate the marginal cost elasticity ε and the time-varying sector-specific vector $\{\kappa_i\}$. For the cost elasticity, we resort to existing estimates suggesting that ε is between one and two (Yashiv, 2007; Merz

[47] Because of constant returns in the sector-specific matching function, it is the convexity of the cost function that prevents concentrating all vacancies and unemployed workers in the sector with the highest efficiency. We follow the convention, common in this literature, that this cost has to be paid every period the vacancy is maintained open.

[48] We could also introduce an aggregate cost-shifter, common across all sectors. Since the results in this section would be unaffected, we omit it to simplify the notation.

and Yashiv, 2007; Coşar, Guner, and Tybout, 2010). Up to this point, we could abstain from modeling behavior and choices of firms and workers in equilibrium. However, measurement of $\{\kappa_i\}$ requires imposing a minimal amount of structure on the equilibrium data generating process. Three assumptions suffice: 1) free entry of vacancies in each sector; 2) no within-market congestion externality, in the spirit of Hosios (1990); and 3) a bargaining protocol between firms and workers such that the firm obtains a share λ, and the worker a share $(1-\lambda)$, of the expected discounted output flow—in particular, outside options do not matter for the bargaining outcome (as in Shaked and Sutton, 1984; Acemoglu, 1996).[49]

This choice of bargaining protocol is convenient because, in the absence of within-sector congestions, it enables us to remain agnostic about the determination of the equilibrium value of unemployment for a worker—therefore reducing to a bare minimum the structure needed on the equilibrium model—and because it isolates mismatch unemployment as the unique source of discrepancy between the efficient and equilibrium distributions of vacancies.[50]

To clearly see this last point, we must compare the equilibrium condition for vacancy creation in sector i to that of the planner. We begin from the equilibrium condition in the economy of Section 2.2 with heterogeneity in $\{\phi_i, z_i, \delta_i, \kappa_i\}$:

$$\kappa_i^{\varepsilon}\left(v_i\right)^{\varepsilon} = \Phi\phi_i\left(\frac{v_i}{u_i}\right)^{\alpha-1}\lambda\frac{Zz_i}{1-\beta\left(1-\Delta\right)\left(1-\delta_i\right)} \tag{14}$$

stating that the marginal cost of a vacancy in sector i (the left hand side), also heterogeneous across sectors, is equated to its expected marginal gain for the firm (the right hand side). Note that the individual firm takes the sectoral meeting probability as given. Note that, as $\varepsilon \to \infty$, $v_i = \kappa_i$, i.e., vacancies are entirely cost-determined. This special case corresponds to the economy of Section 2.

All variables in condition (14) are observable, except for κ_i and ε. For a given value of the elasticity ε, we derive the sequence for κ_i that makes that condition hold

[49]The extensive form game corresponding to this bargaining outcome is spelled out in Acemoglu (1996, Appendix 1). The key assumption is that if, once the pair is formed, a party wants to quit the bargaining, it can rematch within the period within the same sector (i.e., with an identical partner) by paying a small fixed cost.

[50]With the more common Nash bargaining protocol, another discrepancy would arise between the equilibrium value of unemployment and the net shadow value of an additional unemployed worker for the planner $\mu-\xi$, see equation $(A14)$.

exactly at every date in each sector. This strategy amounts to attributing, residually, fluctuations in vacancies to variation in the cost of job creation, once exogenous variation in productivity and separation rates have been accounted for.[51]

7.2 Comparison between equilibrium and planner FOCs

In Appendix A.5, we show that the planner problem of Section 2.2, augmented with a vacancy creation decision where the planner faces the cost function (13), yields the first-order condition

$$\kappa_i^{\varepsilon} (v_i^*)^{\varepsilon} = \Phi \phi_i \left(\frac{v_i^*}{u_i^*}\right)^{\alpha-1} \alpha \frac{Z z_i}{1-\beta(1-\Delta)(1-\delta_i)} \tag{15}$$

equating the marginal cost of a vacancy to its marginal gain, in turn equal to the expected discounted value of output conditional on matching times the marginal decline in the probability of meeting an unemployed worker allocated to sector i.

A comparison of equations (14) and (15) is instructive. With $\lambda = \alpha$ within-market congestion externalities are ruled out and the only reason why equilibrium vacancies in sector i differ from their efficient counterpart is that the number of unemployed workers is the "wrong" one, i.e., the only reason is mismatch unemployment. If in equilibrium an excessive number of unemployed workers search for jobs in declining sectors, firms would create more vacancies than the planner in those sectors, amplifying the initial source of misallocation. Combining equations (14) and (15), and maintaining the assumption $\lambda = \alpha$, we arrive at

$$\frac{v_i}{v_i^*} = \left(\frac{u_i}{u_i^*}\right)^{\frac{1-\alpha}{1-\alpha+\varepsilon}}$$

which demonstrates that the extent to which mismatch unemployment, i.e. deviations of u_i from u_i^*, translate into misallocation of vacancies in equilibrium (i.e., deviation

[51]It is well known that productivity shocks alone are unable to explain fluctuations in vacancies in a matching model with standard parameterization (Shimer, 2005). Investigating the fundamental sources of vacancy fluctuations is beyond the scope of this paper. We limit ourselves to point out that recent papers (e.g., Petrosky-Nadeau, 2009) have emphasized the role of credit shocks and asymmetric information in lending for the observed collapse of job creation during the last recession. In these models, this mechanism works through the free entry condition, precisely as a source of fluctuations in κ_i. A planner subject to the same asymmetric information would face the same fluctuations in κ_i.

of v_i from v_i^*) depends on the value of the elasticity ε. If the marginal cost function is steep (ε high), large differences in the ratio (u_i/u_i^*) and, therefore, in meeting probabilities and expected output gains, translate into small differences in the ratio (v_i/v_i^*) . In this case, planner's vacancies are close to equilibrium vacancies, as assumed in our benchmark analysis. We will provide results for a very broad range of values for ε, and in keeping with the upper bound nature of our exercise, we set $\varepsilon = 1$ in our baseline.

In Appendix A.5, we lay out a simple algorithm to compute the planner's optimal allocation of vacancies across sectors $\{v_{it}^*\}$, and we explain how to modify the calculation of counterfactual unemployment to take into account this additional margin of choice for the planner. It is instructive to examine the relationship between the planner and the equilibrium aggregate job-finding rate in this economy:

$$f_t^* = f_t \cdot \underbrace{\frac{1}{(1-\mathcal{M}_t^x)}}_{\text{Direct Effect}} \cdot \underbrace{\left(\frac{u_t}{u_t^*}\right)^{\alpha}}_{\text{Feedback through } u} \cdot \underbrace{\left[\left(\frac{\bar{\phi}_{xt}^*}{\bar{\phi}_{xt}}\right) \cdot \left(\frac{v_t^*}{v_t}\right)^{\alpha}\right]}_{\text{Feedback through } v}, \tag{16}$$

where $\bar{\phi}_{xt}$ is given by equation (10) and $\bar{\phi}_{xt}^*$ is the same aggregator, but with the planner's vacancy shares v_{it}^*/v_t^* instead of the observed shares. Compared to (11), the equation above features an additional feedback effect of mismatch that operates through vacancies and has two components. Mismatch reduces the aggregate job-finding rate by (i) distorting the distribution of vacancy shares across sectors, and (ii) lowering total vacancies.

7.3 Results

We first estimate the vacancy cost creation parameters κ_i by sector. Next, we compute the distribution of planner's vacancies and the implied planner's aggregate job-finding rate with endogenous vacancies (16), which we then feed into the law of motion for the unemployment rate to perform our counterfactual exercise. Our estimates of the vacancy cost parameter κ increase for almost all industries and occupations during the recession, therefore contributing to the observed drop in vacancies. Figure B19 in Appendix B.3 plots the estimated sequences of κ_i in some selected industries for the case $\varepsilon = 1$.

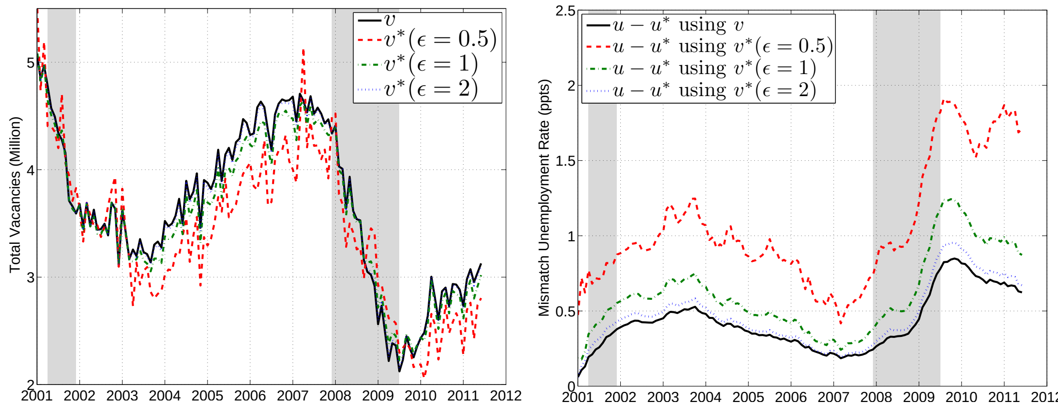


Figure 8: Aggregate vacancies and (left panel) and corresponding mismatch unemployment rates (right panel) at the industry level using endogenous vacancies specification with JOLTS.

Table 1 summarizes the results.[52] We first present our analysis by industry. Figure 8 (left panel) plots aggregate vacancies v_t^* in the planner's economy for different values of ε. The main result is that quantitatively significant deviations between v_t^* and v_t (the data) occur only for low values of the cost elasticity ε. For $\varepsilon \geq 1$, planner and equilibrium vacancies line up closely. This finding is reflected into the calculation of mismatch unemployment (right panel). For $\varepsilon = 1$, with endogenous vacancy creation, mismatch unemployment rises by 0.89 percentage points between 2006 and October 2009, i.e., an additional 0.3 percentage points relative to the exogenous vacancy calculation.[53] For $\varepsilon = 0.5$, mismatch unemployment is generally higher, but its increase between 2006 and October 2009 is still about 1.2 percentage points—not far from the case of unit elasticity.

Turning to occupations, for $\varepsilon = 1$, planner and equilibrium vacancies line up fairly closely and, as Figure 9 indicates, the contribution of mismatch unemployment to the rise in the U.S. unemployment rate between 2006 and October 2009 is 1.1 percentage points.[54] For $\varepsilon = 0.5$, it increases up to 1.5%, or 28% of the total rise in unemployment.

To summarize, as expected, the contribution of mismatch unemployment is larger

[52] The indexes computed with endogenous vacancies have superscript v^*.

[53] Figure B20 in the Appendix also reports our analysis with endogenous vacancies done with the $\mathcal{M}_t$ index. Here, mismatch unemployment rises by about 1.1 percentage points between 2006 and October 2009.

[54] In the case of no heterogeneity in matching and productive efficiency across markets, that contribution rises to 1.8 percentage points, or roughly one third of the total rise in unemployment as shown in Figure B21.

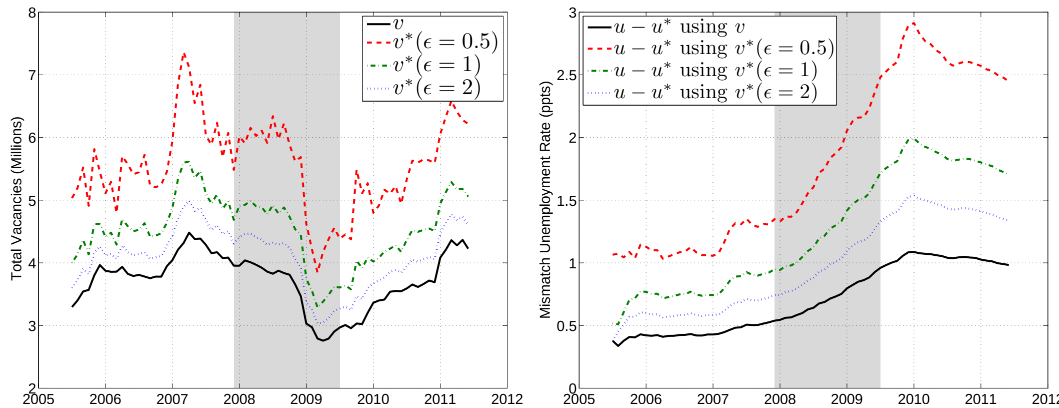


Figure 9: Aggregate vacancies and (left panel) and corresponding mismatch unemployment rates (right panel) at the occupation level using endogenous vacancies specification with the HWOL (The Conference Board Help Wanted OnLine Data Series).

when the distribution of vacancies is endogenized. Nevertheless, our results of Section 5 derived under exogenous vacancies (or infinite marginal cost elasticity) are close to those obtained from the model with endogenous vacancy creation and unitary marginal cost elasticity, a value supported by existing estimates. Our calculations also show that mismatch could have played a major role in the recent rise of unemployment, by dampening aggregate vacancy creation, only if one is willing to maintain that the cost elasticity is very low (below 1/2).

8 Conclusion

How much did mismatch contribute to the dynamics of U.S. unemployment around the Great Recession? To address this question, we developed a framework to coherently define and measure mismatch unemployment. Plausible parameterizations of the model imply that mismatch can explain at most 1/3 of the recent rise in the U.S. unemployment rate. Our formalization of mismatch, and several choices made in our measurement exercise, mean that this estimate should be considered as an upper bound.

While, admittedly, our approach does not put us in the best position to separately identify the many potential causes of mismatch, we argued that analyzing different layers of disaggregation (e.g., occupation, industry, education, geography), as we do, is informative nevertheless. The absence of an increase in geographical mismatch casts doubts on the "house lock" hypothesis, a conclusion in line with existing research. The non-negligible role played by occupational mismatch, especially for high-skilled workers, leaves room for explanations based on labor demand shifts combined with human capital specialization or with relative wage rigidity. Alvarez and Shimer (2010), and Carrillo-Tudela and Visscher (2010), among others, have proposed models where unemployed workers, in equilibrium, make explicit mobility decisions across distinct labor markets. While less amenable to disaggregated measurement than our framework, these models are, potentially, well suited to study the structural causes of mismatch.

If mismatch only accounts for a portion of the persistently high unemployment rate, what are the other economic forces at work? Both the aggregate vacancy rate and aggregate matching efficiency are still below their pre-recession level. Weak aggregate demand combined with wage rigidity (Shimer, 2012), uncertainty about future productivity (Schaal, 2012) and future economic policy (Baker, Bloom, and Davis, 2011), or selective restructuring by firms during recessions (Berger, 2012) do, qualitatively, imply a slow recovery in job creation. The disincentive effects on job search effort from prolonged extension of unemployment benefits (Farber and Valletta, 2011), and the diminished recruitment intensity on firm's side (Davis, Faberman, and Haltiwanger, 2012) are consistent with the fall in aggregate matching efficiency. Going forward, disentangling these channels will be paramount in achieving a comprehensive picture of the Great Recession.

References

[1] Abraham, K. and L.F. Katz (1986). "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?" *Journal of Political Economy*, 94, 507-522.

[2] Acemoglu, D. (1996). "A Microfoundation For Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*, 111, 779-804.

[3] Acemoglu, D. and D. Autor (2011). "Skills, Tasks and Technologies: Implications for Employment and Earnings," in O. Ashenfelter and D. Card (Eds.), Handbook of Labor Economics, Vol. 4B, Chapter 12, pp. 1043-1171. Elsevier.

[4] Alvarez F. and R. Shimer (2010). "Search and Rest Unemployment," *Econometrica*, 79:1, 75-122.

[5] Baker, S. R., N. Bloom, and S. J. Davis (2011). "Measuring Economic Policy Uncertainty," mimeo.

[6] Barlevy, G. (2011). "Evaluating the Role of Labor Market Mismatch in Rising Unemployment," Federal Reserve Bank of Chicago, Economic Perspectives, Vol. 35, Third Quarter.

[7] Barnichon R. and A. Figura (2011). "Labor Market Heterogeneities, Matching Efficiency and the Cyclical Behavior of the Job finding Rate," mimeo.

[8] Berger, D. (2012). "Countercyclical Restructuring and Jobless Recoveries," mimeo.

[9] Borowczyk-Martins, D., G. Jolivet, and F. Postel-Vinay (2012). "Accounting For Endogeneity in Matching Function Estimation," *Review of Economic Dynamics*, Forthcoming.

[10] Carrillo-Tudela C. and L. Visscher (2010). "Search and Endogenous Reallocation over the Business Cycle," mimeo.

[11] Coşar, K. , N. Guner, and J. Tybout (2010). "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy," NBER Working Paper 16326.

[12] Daly, M., B. Hobijn, A. Şahin, and R. Valletta (2011). "A Rising Natural Rate of Unemployment: Transitory or Permanent?," *Journal of Economic Perspectives*, Forthcoming.

[13] Davis, S. J., Faberman, R. J., and J.C. Haltiwanger (2010). "The Establishment-Level Behavior of Vacancies and Hiring," NBER Working Paper 16265.

[14] Davis, S. J., Faberman, R. J., and J.C. Haltiwanger (2012). "Recruiting Intensity during and after the Great Recession: National and Industry Evidence," NBER Working Paper 17782.

[15] Dickens, W. T. (2010). "Has the Recession Increased the NAIRU?", mimeo, Northeastern University and the Brookings Institution.

[16] Eeckhout, J. and P. Kircher (2011). "Identifying Sorting: In Theory," *Review of Economic Studies*, 78(3), 872-906.

[17] Elsby, M., B. Hobijn, and A. Şahin (2010). "The Labor Market in the Great Recession," *Brookings Papers on Economic Activity*, Spring 2010, 1-48.

[18] Faberman, R. J. (2009). "Studying the Labor Market with the Job Openings and Labor Turnover Survey," in Producer Dynamics: New Evidence from Micro Data Volume, edited by T. Dunne, J. Jensen, and M. Roberts, U. of Chicago Press.

[19] Farber H. S. (2012). "Unemployment in the Great Recession: Did the Housing Market Crisis Prevent the Unemployed from Moving to Take Jobs?," *American Economic Review: Papers & Proceedings*, 102(3), 520525

[20] Farber, H. S. and R. Valletta (2011). "Extended Unemployment Insurance and Unemployment Duration in the Great Recession: The U.S. Experience," mimeo.

[21] Galenianos, M. (2012). "Hiring Through Referrals," mimeo.

[22] Hall, R.E. (2010). "The Labor Market in the Current Slump," mimeo.

[23] Herz, B. and T. Van Rens (2011). "Structural Unemployment," mimeo.

[24] Hobijn, B. (2012). " Industry-Occupation Mix," mimeo.

[25] Hosios, A. (1990). "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies* 57, 279-298.

[26] Hsieh, C. and P. Klenow (2009). "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403-1448.

[27] Jackman, R. and S. Roper (1987). "Structural Unemployment," *Oxford Bulletin of Economics and Statistics*, 49, 9-36.

[28] Jaimovich, N. and H. Siu (2012). "The Trend is the Cycle: Job Polarization and Jobless Recoveries," mimeo.

[29] Jones C.I. (2011). "Misallocation, Economic Growth, and Input-Output Economics," mimeo.

[30] Kaplan, G. and S. Schulhofer-Wohl (2010). "Interstate Migration Has Fallen Less Than You Think: Consequences of Hot Deck Imputation in the Current Population Survey," FRB Minneapolis Working Paper 681.

[31] Lagos, R. (2006). "A Model of TFP," *Review of Economic Studies* 73, 983-1007.

[32] Lazear, E. and J. Spletzer (2012). "The United States Labor Market: Status Quo or A New Normal?," mimeo.

[33] Lilien, D.M. (1982). "Sectoral Shifts and Cyclical Unemployment," *Journal of Political Economy*, 90, 777-793.

[34] Merz, M. and E. Yashiv (2007). "Labor and the Market Value of the Firm," *American Economic Review*, 97, 1419-1431.

[35] Mian, A. and A. Sufi (2011). "What Explains High Unemployment? The Aggregate Demand Channel," mimeo.

[36] Mincer, J. (1966). "The Concept of Job Vacancies in a Dynamic Theory of the Labor Market," in The Measurement and Interpretation of Job Vacancies, Columbia University Press.

[37] Moll, B. (2011). "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," mimeo.

[38] Mortensen, D.T. and Nagypal (2007). "More on Unemployment and Vacancy Fluctuations," *Review of Economic Dynamics*, 10, 327-347.

[39] Padoa-Schioppa F. (1991). Mismatch and Labour Mobility. Centre for Economic Policy Research. Cambridge University Press.

[40] Petrosky-Nadeau, N. (2009). "Credit, Vacancies and Unemployment Fluctuations," GSIA Working Paper 2009-E27.

[41] Pissarides C., Equilibrium Unemployment Theory (2nd edition), MIT Press, Cambridge, MA, 2000.

[42] Restuccia, D. and R. Rogerson (2008). "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11, 707-720.

[43] Şahin, A., J. Song, G. Topa, and G.L. Violante (2012). "Mismatch Unemployment," NBER Working Paper 18265.

[44] Schaal, E. (2012). "Uncertainty, Productivity and Unemployment in the Great Recession," mimeo.

[45] Schulhofer-Wohl, S. (2010). "Negative Equity Does Not Reduce Homeowners' Mobility," mimeo.

[46] Shaked, A. and J. Sutton (1984). "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model," *Econometrica*, 52, 13511364.

[47] Shimer R. (2005). "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25-49.

[48] Shimer R. (2007). "Mismatch," *American Economic Review*, 97, 1074-1101.

[49] Shimer R. (2012). "Wage Rigidities and Jobless Recoveries," mimeo.

[50] Veracierto (2011). " Worker Flows and Matching Efficiency," Federal Reserve Bank of Chicago, Economic Perspectives, Vol. 35, Fourth Quarter.

[51] Yashiv, E. (2007). "Labor Search and Matching in Macroeconomics," *European Economic Review*, 51, 1859-1895.

A Theoretical Appendix

This Appendix formally derives all the theoretical results of Sections 2 and 7.

A.1 Heterogeneous matching efficiencies

We solve the planner's problem of Section 2.1. The efficient allocation at any given date is the solution of the following planner's problem that we write in recursive form:

$$\begin{aligned}
V(\mathbf{e};\mathbf{v},\phi,Z,\Delta,\Phi) &= \max_{\{u_i\geq 0\}} \sum_{i=1}^{I} Z(e_i+h_i) + \beta\mathbb{E}\left[V(\mathbf{e}';\mathbf{v}',\phi',Z',\Delta',\Phi')\right] \\
s.t. &: \\
\sum_{i=1}^{I}(e_i+u_i) &= 1 && \text{(A1)} \\
h_i &= \Phi\phi_i m(u_i,v_i) && \text{(A2)} \\
e_i' &= (1-\Delta)(e_i+h_i) && \text{(A3)} \\
&\Gamma_{Z,\Delta,\Phi}(Z',\Delta',\Phi';Z,\Delta,\Phi),\ \Gamma_{\mathbf{v}}(\mathbf{v}';\mathbf{v},Z',\Delta',\Phi'),\Gamma_{\phi}(\phi';\phi) && \text{(A4)}
\end{aligned}$$

The per period output for the planner is equal to $Z(e_i+h_i)$ in each market i. The first constraint $(A1)$ states that the planner has $1-\sum_{i=1}^{I} e_i$ unemployed workers available to allocate across sectors. Equation $(A2)$ states that, once the allocation $\{u_i\}$ is chosen, the frictional matching process in each market yields $\Phi\phi_i m(u_i,v_i)$ new hires which add to the existing e_i active matches. Equation $(A3)$ describes separations and the determination of next period's distribution of active matches $\{e_i'\}$ in all sectors. Line $(A4)$ in the problem collects all the exogenous stochastic processes the planner takes as given.

It is easy to see that this is a concave problem where first-order conditions are sufficient for optimality. At an interior solution ($u_i>0$ for all i), the choice of how many unemployed workers u_i to allocate in market i yields the first-order condition

$$Z\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \beta\mathbb{E}\left[V_{e_i}(\mathbf{e}';\mathbf{v}',\phi',Z',\Delta',\Phi')\right](1-\Delta)\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu, \tag{A5}$$

where μ is the multiplier on constraint $(A1)$. The right-hand side (RHS) of this condition is the shadow value of an additional worker in the unemployment pool available to search. The left-hand side (LHS) is the expected marginal value of an additional unemployed worker allocated to sector i. The derivative of the sector-specific matching function m is written as a function of local market tightness only (with a slight abuse of notation) because of its CRS specification.

The Envelope condition with respect to the state e_i yields:

$$V_{e_i}(\mathbf{e};\mathbf{v},\phi,Z,\Delta,\Phi) = Z - \mu + \beta(1-\Delta)\mathbb{E}\left[V_{e_i}(\mathbf{e}';\mathbf{v}',\phi',Z',\Delta',\Phi')\right], \qquad \text{(A6)}$$

from which it is immediate to see, by iterating forward, that $\mathbb{E}\left[V_{e_i}(\mathbf{e}';\mathbf{v}',\phi',Z',\Delta',\Phi')\right]$ is independent of i, since productivity and the job destruction rate are common across all sectors.[55] Using this result into $(A5)$, the optimal rule for the allocation of unemployed workers across sectors can be written as equation (1) in the main text.

A.2 Heterogenous productivities and destruction rates

We extend the baseline model of Section 2.1 as follows. Individuals (still in measure one) can be either employed in sector i (e_i), or unemployed and searching in sector i (u_i), or out of the labor force. The aggregate labor force is $\ell = \sum_{i=1}^{I}(e_i+u_i) \leq 1$. We normalize to zero utility from non participation, and let $\xi > 0$ denote the disutility of search for the unemployed.

Labor productivity in sector i is given by $Z \cdot z_i$, where each idiosyncratic component z_i is strictly positive, i.i.d. across sectors and independent of Z. Let the conditional distribution of the vector $\mathbf{z} = \{z_i\}$ be $\Gamma_z(\mathbf{z}',\mathbf{z})$. The idiosyncratic component of the exogenous destruction rate in sector i is δ_i, i.i.d. across sectors and independent of Δ, Z and z_i. Let the conditional distribution of the vector $\delta = \{\delta_i\}$ be $\Gamma_\delta(\delta',\delta)$. The survival probability of a match is then $(1-\Delta)(1-\delta_i)$. The vector $\{Z,\Delta,\Phi,\mathbf{z},\mathbf{v},\phi,\delta\}$ takes strictly positive values.

It is convenient to impose additional structure on some conditional distributions: as specified in the text, we assume that $(Z, 1-\Delta, z_i, 1-\delta_i)$ are all positive martingales. The timing of events is exactly as before, with the decision on the size of the labor force for next period taken at the end of the current period. The recursive formulation of the planner's problem has three additional states compared to the problem of Section 2.1: the current number of unemployed workers u, the vector of productive efficiencies $\mathbf{z}$, and the vector of destruction rates δ. The

[55] We are also using the transversality condition $\lim_{t\to\infty}\beta^t(1-\Delta)^t\mathbb{E}\left[V_{e_{it}}\right] = 0$.

planner solves the problem:

$$\begin{aligned}
V\left(u,\mathbf{e};\mathbf{z},\mathbf{v},\phi,\delta,Z,\Delta,\Phi\right) &= \max_{\{u_i,\ell'\}}\sum_{i=1}^{I} Zz_i\left(e_i+h_i\right)-\xi u+\beta\mathbb{E}\left[V\left(u',\mathbf{e}';\mathbf{z}',\mathbf{v}',\phi',\delta',Z',\Delta',\Phi'\right)\right]\\
s.t. &:\\
\sum_{i=1}^{I} u_i &\leq u \qquad \text{(A7)}\\
h_i &= \Phi\phi_i m\left(u_i,v_i\right) \qquad \text{(A8)}\\
e_i' &= \left(1-\Delta\right)\left(1-\delta_i\right)\left(e_i+h_i\right) \qquad \text{(A9)}\\
u' &= \ell'-\sum_{i=1}^{I} e_i' \qquad \text{(A10)}\\
u_i &\in \left[0,u\right],\ell'\in\left[0,1\right], \qquad \text{(A11)}
\end{aligned}$$

$$\Gamma_{Z,\Delta,\Phi}\left(Z',\Delta',\Phi';Z,\Delta,\Phi\right),\ \Gamma_{\mathbf{v}}\left(\mathbf{v}';\mathbf{v},Z',\Delta',\Phi',\mathbf{z}'\right),\Gamma_{\phi}\left(\phi';\phi\right),\Gamma_{\mathbf{z}}\left(\mathbf{z}';\mathbf{z}\right),\Gamma_{\delta}\left(\delta',\delta\right) \qquad \text{(A12)}$$

where the conditional distributions in the last line are restricted as described above. The choice of how many unemployed workers u_i to allocate in the i market yields the first-order condition

$$Zz_i\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right)+\beta\mathbb{E}\left[-V_u'\left(\cdot\right)+V_{e_i}'\left(\cdot\right)\right]\left(1-\Delta\right)\left(1-\delta_i\right)\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right)=\mu, \qquad \text{(A13)}$$

where μ is the multiplier on constraint $(A7)$. The Envelope conditions with respect to the states u and e_i yield:

$$\begin{aligned}
V_u\left(u,\mathbf{e};\mathbf{z},\mathbf{v},\phi,\delta,Z,\Delta,\Phi\right) &= \mu-\xi \qquad \text{(A14)}\\
V_{e_i}\left(u,\mathbf{e};\mathbf{z},\mathbf{v},\phi,\delta,Z,\Delta,\Phi\right) &= Zz_i+\beta(1-\Delta)\left(1-\delta_i\right)\mathbb{E}\left[V_{e_i}'-V_u'\right]. \qquad \text{(A15)}
\end{aligned}$$

According to the first Envelope condition, the marginal value of an unemployed to the planner equals the shadow value of being available to search (μ) net of the disutility of search ξ. The second condition states that the marginal value of an employed worker is its flow output this period plus its discounted continuation value net of the value of search, conditional on the match not being destroyed.

The optimal decision on the labor force size next period ℓ' requires

$$\mathbb{E}\left[V_u\left(u',\mathbf{e}';\mathbf{z}',\mathbf{v}',\phi',\delta',Z',\Delta',\Phi'\right)\right]=0, \qquad \text{(A16)}$$

i.e., the expected marginal value of moving a nonparticipant into job search should be equal to its value as nonparticipant, normalized to zero. By combining $(A16)$ with $(A14)$, we note that the planner will choose the size of the labor force so that the expected shadow value of an unemployed worker $\mathbb{E}\left[\mu'\right]$ equals search disutility ξ.[56]

Using $(A16)$ into the Envelope condition $(A15)$ under the additional assumption that all the elements of the vector $x = (Z, 1-\Delta, z_i, 1-\delta_i)$ are independent martingales, and iterating forward, we arrive at:

$$\mathbb{E}\left[V'_{e_i}\right] = \frac{Zz_i}{1-\beta\left(1-\Delta\right)\left(1-\delta_i\right)} \tag{A17}$$

which, substituted into equation $(A13)$ yields

$$Zz_i\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \frac{\beta\left(1-\Delta\right)\left(1-\delta_i\right)}{1-\beta\left(1-\Delta\right)\left(1-\delta_i\right)}Zz_i\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu. \tag{A18}$$

Rearranging, we conclude that the planner allocates idle labor to equalize

$$\frac{z_i}{1-\beta\left(1-\Delta\right)\left(1-\delta_i\right)}\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right)$$

across sectors, which is expression (2) in Section (2.2) in the main text.

A.3 Endogenous separations

Consider the environment of Section 2.2 and allow the planner to move workers employed in sector i into unemployment (or out of the labor force) at the end of the period, before choosing the size of the labor force for next period. There are two changes to the planner's problem. First, the law of motion for employment becomes

$$e'_i = \left(1-\Delta\right)\left(1-\delta_i\right)\left(e_i+h_i\right) - \sigma_i. \tag{A19}$$

Second, the planner has another vector of choice variables $\{\sigma_i\}$, with $\sigma_i \in \left[0, \left(1-\Delta\right)\left(1-\delta_i\right)\left(e_i+h_i\right)\right]$.

The decision of how many workers to separate from sector i employment into unemployment is:

$$\mathbb{E}\left[V'_u\left(\cdot\right) - V'_{e_i}\left(\cdot\right)\right]\begin{cases} <0 & \to \sigma_i = 0 \\ =0 & \to \sigma_i \in \left(0, \left(1-\Delta\right)\left(1-\delta_i\right)\left(e_i+h_i\right)\right) \\ >0 & \to \sigma_i = \left(1-\Delta\right)\left(1-\delta_i\right)\left(e_i+h_i\right) \end{cases} \tag{A20}$$

depending on whether at the optimum a corner or interior solution arises. If the first-order conditions $(A16)$ hold with equality, then the optimality condition $(A20)$ holds with the "$<$" inequality and $\sigma_i = 0$. As a result, the planner's allocation rule (2) remains unchanged.

[56] It is clear that our result is robust to allowing ξ to be stochastic and correlated with (Z, Δ, Φ).

A.4 Properties of the mismatch index

First, we prove that $0 \leq \mathcal{M}_{\phi t} \leq 1$. Since all the components of the sum in (7) are positive, $\mathcal{M}_{\phi t} \leq 1$. Under maximal mismatch (no markets where unemployment and vacancies coexist), the index is exactly equal to one. To show that $\mathcal{M}_{\phi t} \geq 0$, note that

$$
\begin{aligned}
1 - \mathcal{M}_{\phi t} &= \frac{1}{v_t^{\alpha} u_t^{1-\alpha}} \frac{1}{\left[\sum_{i=1}^{I} \phi_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t}\right)\right]^{\alpha}} \sum_{i=1}^{I} \left(\phi_{it}^{\frac{1}{\alpha}} v_{it}\right)^{\alpha} (u_{it})^{1-\alpha} \\
&\leq \frac{1}{v_t^{\alpha} u_t^{1-\alpha}} \frac{1}{\left[\sum_{i=1}^{I} \phi_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t}\right)\right]^{\alpha}} \left[\sum_{i=1}^{I} \left(\phi_{it}^{\frac{1}{\alpha}} v_{it}\right)\right]^{\alpha} \left(\sum_{i=1}^{I} u_{it}\right)^{1-\alpha} \\
&= 1
\end{aligned}
$$

where the $\leq$ sign follows from Hölder's inequality. It is easy to show that the index becomes exactly zero in absence of mismatch by substituting the allocation rule (5) into the index.

By inspecting (7) , it is also easy to see that the $\mathcal{M}_{\phi t}$ index is invariant to "pure" aggregate shocks that shift the total number of vacancies and unemployed up or down, but leave the vacancy and unemployment shares across markets unchanged.

To show that the mismatch index is increasing in the level of disaggregation, consider an economy where the aggregate labor market is described by two dimensions indexed by (i, j), e.g., I regions $\times$ J occupations. Let $\mathcal{M}_{\phi It}$ be the mismatch index over the I sectors and $\mathcal{M}_{\phi IJt}$ be the one over the $I \times J$ sectors. From the disaggregated matching function, we have $h_{ijt} = \Phi \phi_{ijt} v_{ijt}^{\alpha} u_{ijt}^{1-\alpha}$. Summing this expression over j yields

$$
h_{it} = \sum_{j=1}^{J} \Phi \phi_{ijt} v_{ijt}^{\alpha} u_{ijt}^{1-\alpha} = \Phi \left[\sum_{j=1}^{J} \phi_{ijt} \left(\frac{v_{ijt}}{v_{it}}\right)^{\alpha} \left(\frac{u_{ijt}}{u_{it}}\right)^{1-\alpha}\right] v_{it}^{\alpha} u_{it}^{1-\alpha}. \tag{A21}
$$

At the aggregated level, we have $h_{it} = \Phi \phi_{it} v_{it}^{\alpha} u_{it}^{1-\alpha}$ and so (A21) implies that

$$
\phi_{it} = \sum_{j=1}^{J} \phi_{ijt} \left(\frac{v_{ijt}}{v_{it}}\right)^{\alpha} \left(\frac{u_{ijt}}{u_{it}}\right)^{1-\alpha}. \tag{A22}
$$

Now consider the disaggregated matching index. We have

$$
1 - \mathcal{M}_{\phi IJt} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\phi_{ijt}}{\bar{\phi}_{IJt}} \left(\frac{v_{ijt}}{v_t}\right)^{\alpha} \left(\frac{u_{ijt}}{u_t}\right)^{1-\alpha} \tag{A23}
$$

for

$$\bar{\phi}_{IJt} = \left[\sum_{i=1}^{I} \sum_{j=1}^{J} \phi_{ijt}^{\frac{1}{\alpha}} \left(\frac{v_{ijt}}{v_t} \right) \right]^{\alpha}. \tag{A24}$$

Manipulating the above expression yields

$$\begin{aligned} 1 - \mathcal{M}_{\phi IJt} &= \frac{1}{\bar{\phi}_{IJt} v_t^{\alpha} u_t^{1-\alpha}} \sum_{i=1}^{I} \sum_{j=1}^{J} \phi_{ijt} v_{ijt}^{\alpha} u_{ijt}^{1-\alpha} \\ &= \frac{1}{\bar{\phi}_{IJt} v_t^{\alpha} u_t^{1-\alpha}} \sum_{i=1}^{I} v_{it}^{\alpha} u_{it}^{1-\alpha} \sum_{j=1}^{J} \phi_{ijt} \left(\frac{v_{ijt}}{v_{it}} \right)^{\alpha} \left(\frac{u_{ijt}}{u_{it}} \right)^{1-\alpha} \\ &= \frac{1}{\bar{\phi}_{IJt}} \sum_{i=1}^{I} \phi_{it} \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} \end{aligned}$$

where the third step above follows from (A22). Next, manipulating (A24) delivers

$$\begin{aligned} \bar{\phi}_{IJt} &= \left\{ \frac{1}{v_t} \sum_{i=1}^{I} v_{it} \left(\left[\sum_{j=1}^{J} \phi_{ijt}^{\frac{1}{\alpha}} \left(\frac{v_{ijt}}{v_{it}} \right) \right]^{\alpha} \right)^{\frac{1}{\alpha}} \right\}^{\alpha} \\ &= \left\{ \frac{1}{v_t} \sum_{i=1}^{I} v_{it} \left(\left[\sum_{j=1}^{J} \phi_{ijt}^{\frac{1}{\alpha}} \left(\frac{v_{ijt}}{v_{it}} \right) \right]^{\alpha} \cdot \left[\sum_{j=1}^{J} \frac{u_{ijt}}{u_{it}} \right]^{1-\alpha} \right)^{\frac{1}{\alpha}} \right\}^{\alpha} \end{aligned}$$

where the second step above follows from the identity $\sum_{j=1}^{J} u_{ijt} = u_{it}$. Applying Holder's inequality now yields

$$\begin{aligned} \bar{\phi}_{IJt} &\geq \left\{ \frac{1}{v_t} \sum_{i=1}^{I} v_{it} \left(\sum_{j=1}^{J} \phi_{ijt} \left(\frac{v_{ijt}}{v_{it}} \right)^{\alpha} \left(\frac{u_{ijt}}{u_{it}} \right)^{1-\alpha} \right)^{\frac{1}{\alpha}} \right\}^{\alpha} \\ &= \left\{ \sum_{i=1}^{I} \phi_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t} \right) \right\}^{\alpha} = \bar{\phi}_{It} \end{aligned}$$

where $\bar{\phi}_{It}$ is the equivalent expression to $\bar{\phi}_{IJt}$ in (A23) for the aggregated case. Combining results, we have shown that

$$1 - \mathcal{M}_{\phi IJt} \leq \sum_{i=1}^{I} \frac{\phi_{it}}{\bar{\phi}_{It}} \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} = 1 - \mathcal{M}_{\phi It}$$

and so we must have $\mathcal{M}_{\phi IJt} \geq \mathcal{M}_{\phi It}$.

A.5 Planner's problem with endogenous vacancies

Optimal vacancy creation Consider the planner's problem of Section 2.2 solved in Appendix A.2, the most general of our environments, and let the creation of vacancies $\{v_i\}$ be under the control of the planner.

$$
\begin{aligned}
V(u, \mathbf{e}; \mathbf{z}, \phi, \delta, \kappa, Z, \Delta, \Phi) &= \max_{\{u_i, v_i, \ell'\}} \sum_{i=1}^{I} Z z_i (e_i + h_i) - K_i(v_i) - \xi u + \beta \mathbb{E}\left[V(u', \mathbf{e}'; \mathbf{z}', \phi', \delta', \kappa', Z', \Delta', \Phi')\right] \\
s.t. &: \\
\sum_{i=1}^{I} u_i &\leq u \quad \text{(A25)} \\
h_i &= \Phi \phi_i m(u_i, v_i) \quad \text{(A26)} \\
e_i' &= (1-\Delta)(1-\delta_i)(e_i + h_i) \quad \text{(A27)} \\
u' &= \ell' - \sum_{i=1}^{I} e_i' \quad \text{(A28)} \\
u_i &\in [0, u], \ell' \in [0,1], v_i \geq 0 \quad \text{(A29)} \\
&\Gamma_{Z,\Delta,\Phi}(Z', \Delta', \Phi'; Z, \Delta, \Phi),\ \Gamma_\phi(\phi'; \phi), \Gamma_{\mathbf{z}}(\mathbf{z}'; \mathbf{z}), \Gamma_\delta(\delta', \delta), \Gamma_\kappa(\kappa', \kappa) \quad \text{(A30)}
\end{aligned}
$$

The optimality condition for vacancy creation is

$$
K_{v_i}(v_i) = \Phi \phi_i m_{v_i} \left(\frac{v_i}{u_i}\right) \left\{ Z z_i + \beta (1-\Delta)(1-\delta_i) \mathbb{E}\left[V_{e_i}'(\cdot)\right] \right\}.
$$

Using the expression for $\mathbb{E}\left[V_{e_i}'(\cdot)\right]$ obtained in $(A17)$ and the functional forms for K_i and m specified in the main text, we obtain expression (15).

Calculation of planner's vacancies We now lay out an algorithm to compute the planner's optimal allocation of vacancies across sectors. Rearranging condition $(A18)$ dictating the optimal allocation of unemployed workers across sectors, given the distribution of vacancies $\{v_i^*\}$, yields

$$
\frac{v_i^*}{u_i^*} = \left[\frac{\mu}{1-\alpha} \cdot \frac{1}{\frac{Z z_i \Phi \phi_i}{1-\beta(1-\Delta)(1-\delta_i)}}\right]^{\frac{1}{\alpha}} \quad \text{(A31)}
$$

where μ is the multiplier on the resource constraint $\sum_{i=1}^{I} u_i \leq u$. Substituting $(A31)$ into (15) yields an equation for the optimal number of vacancies in sector i which reads

$$
v_i^* = \frac{1}{\kappa_i} \left(\frac{\alpha}{1-\alpha}\right)^{1/\varepsilon} \cdot \left(\frac{1}{\mu}\right)^{\frac{(1-\alpha)/\varepsilon}{\alpha}} \cdot \left[(1-\alpha) \cdot \frac{Z z_i \Phi \phi_i}{1-\beta(1-\Delta)(1-\delta_i)}\right]^{\frac{1/\varepsilon}{\alpha}}. \quad \text{(A32)}
$$

Summing over all $i's$, we arrive at the optimal share of vacancies in sector i

$$\frac{v_i^*}{v_t^*} = \frac{\frac{1}{\kappa_i}\left[\frac{z_i\phi_i}{1-\beta(1-\Delta)(1-\delta_i)}\right]^{\frac{1/\varepsilon}{\alpha}}}{\sum_{i=1}^{I}\frac{1}{\kappa_i}\left[\frac{z_i\phi_i}{1-\beta(1-\Delta)(1-\delta_i)}\right]^{\frac{1/\varepsilon}{\alpha}}} \tag{A33}$$

only as a function of parameters, which is quite intuitive: the higher is productive, matching and job creation efficiency in sector i, relative to the other sectors, the larger its share of vacancies. However, to solve the model, we need to determine the *level* of v_i^* which requires eliminating μ from $(A32)$. Combining again the two first order conditions, and summing across all sectors, we arrive at

$$u^* = \left(\frac{\alpha}{1-\alpha}\right)^{1/\varepsilon}\cdot\left[Z\Phi\cdot(1-\alpha)\right]^{\frac{1+1/\varepsilon}{\alpha}}\cdot\left(\frac{1}{\mu}\right)^{\frac{1+(1-\alpha)/\varepsilon}{\alpha}}\cdot\sum_{i=1}^{I}\frac{1}{\kappa_i}\left[\frac{z_i\phi_i}{1-\beta\left(1-\Delta\right)\left(1-\delta_i\right)}\right]^{\frac{1+1/\varepsilon}{\alpha}}, \tag{A34}$$

which establishes a unique inverse relationship between μ and u^*: the higher the number of idle workers, the lower the shadow value of the constraint.

Equation $(A34)$ suggests an algorithm to solve for v_i^*. At any date, before choosing how to allocate vacancies and unemployed workers, the total number of idle workers is a state variable for the planner, i.e., u^* is known. One can therefore back out μ from $(A34)$, and then v_i^* from $(A32)$ and u_i^* from $(A31)$.

Counterfactual unemployment To perform the counterfactual on unemployment with endogenous vacancies, we use the same iterative procedure described in Section 3.2, with the caveat that the relationship between the planner's job-finding rate and the empirical job-finding rate at date t is now given by

$$f_t^* = \frac{h_t^*}{u_t^*} = \Phi_t\bar{\phi}_{xt}^*\left(\frac{v_t^*}{u_t^*}\right)^{\alpha} = f_t\cdot\frac{1}{1-\mathcal{M}_{xt}}\cdot\left(\frac{u_t}{u_t^*}\right)^{\alpha}\cdot\left[\left(\frac{\bar{\phi}_{xt}^*}{\bar{\phi}_{xt}}\right)\cdot\left(\frac{v_t^*}{v_t}\right)^{\alpha}\right], \tag{A35}$$

where $\bar{\phi}_{xt}$ is given by equation (10), and $\bar{\phi}_{xt}^*$ is the same aggregator with shares (v_{it}^*/v_t^*) instead of (v_{it}/v_t). When $v_{it}^* = v_{it}$ (i.e., $\varepsilon\to\infty$), equation $(A35)$ collapses to the relationship $f^* = \left[f/\left(1-\mathcal{M}_{xt}\right)\right]\left(u_t/u_t^*\right)^{\alpha}$ that we have used in our baseline calculations with exogenous vacancies.

B Data Appendix

B.1 Comparison between JOLTS and HWOL vacancies

Vacancies recorded in JOLTS are derived from a sample of about 16,000 business establishments. JOLTS vacancies represent "all unfilled, posted positions available at an establishment on the last day of the month. The vacancy must be for a specific position where work can start within thirty days, and an active recruiting process must be underway for the position." (Faberman, 2009, p. 86). As noted in Section 4, the HWOL database collects ads from job listings posted by employers on thousands of internet job boards and online newspapers. The HWOL program uses a mid-month survey reference period. For example, data for October would be the sum of all posted ads from September 14th through October 13th. This reference period is aligned to the BLS unemployment "job search" time period. The monthly vacancy counts that we use in our calculations are total monthly unduplicated ads appearing in the reference period. This figure therefore includes both newly posted ads and ads reposted from the previous months.

Sample establishments in the JOLTS only report their own direct employees and exclude "employees of temporary help agencies, employee leasing companies, outside contractors, and consultants," which are counted by their employer of record, not by the establishment where they are working.[57] Thus, this approach captures temp-help and leasing workers as long as their employers are sampled in the JOLTS, but does not capture the self-employed contract workforce (these workers typically receive a 1099-MISC form instead of a W-2 form to report payments received for services they provide). On the other hand, the HWOL series includes postings for contract work. In what follows, we often also report HWOL vacancy counts excluding contract work, to make the series more comparable to the JOLTS measure of vacancies, but in our empirical analyses of mismatch with HWOL data we consider all ads, including those for contract work.

We perform two exercises to compare the vacancy counts we get from each data source, one at the regional level and one at the industry level—region and industry are the only dimensions available in both JOLTS and HWOL. First, we compare total vacancies by Census region in Figure B1. The HWOL series tend to be lower than the JOLTS series before 2008 (especially in the South), and higher from 2008 onwards (especially in the Northeast). The two series are closest in the West: here the correlation between the HWOL and JOLTS series is 0.94. In the other three regions the correlation is lower: 0.27 in the Midwest, 0.40 in the South, and 0.54 in the Northeast. Our re-weighing strategy in Section 6 enables us to correct for the possibility that online ads penetration may differ across regions.

[57] See the JOLTS Technical Note at http://www.bls.gov/news.release/jolts.tn.htm.

For about 57% of the job listings, we observe the NAICS code of the employer. Therefore, we are able to directly compare vacancy counts by industry from HWOL to those in the JOLTS. We report in Figure B2 scatterplots of vacancy shares by industry from JOLTS and from HWOL—for the latter, we report both total vacancies, as well as vacancies without contract work. The top panel of the figure reports average vacancy shares over the sample period under consideration. Most data points are close to the 45-degree line, indicating that the vacancy shares by industry in the two series line up fairly well, especially when we omit contract work from HWOL to make it more comparable to the JOLTS. The only two sectors where JOLTS and HWOL show significant differences in vacancy shares are "Public Administration" and "Accommodation and Food Services." The bottom panel reports the change in average vacancy shares between 2006 and the 12 month period around December 2009 for each series. Again, the JOLTS and HWOL series are quite close to each other, with the exception of "Public Administration."

We have investigated whether the missing industry information in HWOL exhibits any systematic patterns that may have skewed our analysis. For robustness, we re-weighted the industry observations in HWOL as follows: first, we dropped observations from individual Job Boards with the highest rates of missing NAICS codes. Then, we re-weighted the remaining observations to correct for any correlation between NAICS missing values and Job Board, occupation or Census region. In other words, if vacancies for specific (Job Board, SOC, Census region) combinations are more likely to have missing NAICS codes, the vacancies that do have NAICS information in those cells are assigned a larger weight in computing total vacancies by industry.[58] The resulting vacancy shares are almost identical to those based on the raw data.

To sum up, the comparison between JOLTS and HWOL vacancy counts suggests that there are some discrepancies in the behavior of two series. The main concerns are (i) the possible over- or under-use of online advertisement in certain sectors (regions and/or industries) and (ii) the presence of an upward trend in the use of online recruitment that could artificially mitigate the drop in job advertisements around the last recession (and inflate the subsequent recovery). We address these issues in Section 6 and show that our quantitative results on mismatch measures are robust.

[58]For example, suppose a (Job Board, SOC, Census region) cell has four observations. Observation one is in NAICS code 11, observations two and three are in NAICS code 13, and observation four has a missing NAICS. Thus, the missing NAICS rate is 0.25. Then, a weight of $1/(1-0.25) = 1.333$ is applied to each observation with non-missing NAICS. So we find 1.333 job vacancies in NAICS code 11, and 2.667 job vacancies in NAICS code 13.

B.2 Matching function estimation

Throughout our analysis we assume matching functions are constant returns to scale. We begin by imposing a Cobb-Douglas specification. At the end of this section we show that, when we allow for a more general CES specification, our results point towards an elasticity of substitution statistically close to one.

To compute market-specific matching efficiency parameters, ϕ_i, and vacancy share α, we use various data sources. At the industry level, we use vacancies and hires from JOLTS, and unemployment counts from the CPS. At the occupation level, we use vacancies from HWOL but do not have a direct measure of hires as in JOLTS. Therefore, we construct hires from the CPS using flows from unemployment into a given occupation i for people who are surveyed in adjacent months. Because these monthly flows are quite noisy, we use a 3-month moving average of the data, and aggregate occupations into five broad occupation groups. For comparison purposes, we replicate the analysis at the industry level using the constructed "CPS hires" as well. At the aggregate level, we perform the estimation using both JOLTS and HWOL vacancies, and both JOLTS and CPS hires.

The estimation of matching functions is subject to an endogeneity problem, as shocks to unobserved matching efficiency may affect the number of vacancies posted by firms—much like TFP shocks affect firm's choice of labor input. To deal with this issue, we follow two strategies suggested by Borowczyk-Martins, Jolivet, and Postel-Vinay (2012). First, they recognize that some of the major movements in matching efficiency inducing a bias in the OLS estimator are low-frequency ones. As a result, modeling explicitly the dynamics of matching efficiency through time-varying polynomials and structural breaks goes a long way towards solving the problem even with the simple OLS estimator. This is the first route we take. At the aggregate level, we estimate:

$$\log\left(\frac{h_t}{u_t}\right) = const + \gamma' QTT_t + \alpha \log\left(\frac{v_t}{u_t}\right) + \epsilon_t, \tag{B1}$$

where QTT_t is a vector of four elements for the quartic time trend which is meant to capture shifts in aggregate matching efficiency.

At the sectoral level, we are interested in the sector-specific component of matching efficiency orthogonal to common aggregate movements in aggregate matching efficiency. Therefore, at the industry and 2-digit occupation level, we perform the following panel regression:

$$\log\left(\frac{h_{it}}{u_{it}}\right) = \gamma' QTT_t + \chi_{\{t\leq 07\}} \log\left(\phi_i^{pre}\right) + \chi_{\{t>07\}} \log\left(\phi_i^{post}\right) + \alpha \log\left(\frac{v_{it}}{u_{it}}\right) + \epsilon_{it}, \tag{B2}$$

where $\chi_{\{t>07\}}$ is an indicator for months after December 2007, the official start of the recession, to absorb sector-specific shifts in matching efficiency.

Borowczyk-Martins, Jolivet and Postel-Vinay (2012) also propose a GMM estimator to take care of the simultaneity bias. This method requires imposing an ARMA(p,q) structure on the matching efficiency process: we follow their model selection protocol and set $p = 3$ and $q = 3$. We use an over-identified GMM estimator implemented with four lags of market tightness and one lag of the job-finding rate as instruments, as they argue it is the one delivering the most precise parameter estimates.

Table B4 displays the full set of estimates of the vacancy share parameter α. In the aggregate regressions, the estimated vacancy share varies between 0.32 and 0.67; in the panel regressions, the estimates are somewhat lower varying between 0.24 and 0.53. To construct our mismatch indices, and in our calculation of mismatch unemployment, we pick a value of $\alpha = 0.5$ for two reasons. First, it is the midpoint of our estimates with aggregate data. Second, our mismatch indices are typically highest for $\alpha = 0.5$; therefore, in the spirit of reporting an upper bound for mismatch unemployment, we use this value.

The estimated quartic time trend (not shown) drops during the recession in all our OLS specifications, consistent with a deterioration of aggregate matching efficiency. With regard to sectoral matching efficiency, in our baseline calculations we use the estimates obtained with JOLTS hires for the industry level mismatch analysis, and those with CPS hires for the occupation level analysis. In all cases, we use the *pre-recession* matching efficiency parameter estimates, and verify the robustness of our findings to this choice. The estimated matching efficiency parameters ϕ_i pre- and post-recession are reported in Appendix B, Tables B6-B8. Beyond movements in the common component Φ_t, changes over time in sector-specific matching efficiencies are small.

Finally, in order to examine the plausibility of the Cobb-Douglas specification, we generalize (B2) and estimate the following CES specification via minimum distance:

$$\log\left(\frac{h_{it}}{u_{it}}\right) = \gamma' QTT_t + \chi_{\{t\leq 07\}} \log\left(\phi_i^{pre}\right) + \chi_{\{t>07\}} \log\left(\phi_i^{post}\right) + \frac{1}{\sigma}\log\left[\alpha\left(\frac{v_{it}}{u_{it}}\right)^{\sigma} + (1-\alpha)\right] + \epsilon_{it}. \tag{B3}$$

Recall that $\sigma \in (-\infty, 1)$ with $\sigma = 0$ in the Cobb-Douglas case. A simulated annealing algorithm is used to ensure that we obtain a global minimum. 95% confidence intervals are computed via bootstrap methods. The estimation results are reported in Table B5. The point estimates of σ range from -0.11 to 0.18 depending on the specification, implying an elasticity between 0.9 and 1.2. In the specification with HWOL vacancies and CPS hires, we cannot reject the null that $\sigma = 0$ at the 5% significance level. In the other specifications with JOLTS data, $\sigma = 0$ lies just outside the 95% confidence interval, but the point estimates are close to zero, implying values close to unity for the elasticity of the matching function.

B.3 Additional results on industrial and occupational mismatch

To examine the robustness of our results, we present a number of additional specifications. Table B9 summarizes the results. We first compute the indexes adjusted for one source of heterogeneity at a time (see Figure B6). Next, we compute the indexes for different values of α (Figure B7). In addition, we compute the mismatch index using ϕ's separately estimated for the periods before and after the recession. We denote this index as $\mathcal{M}_x^{break}$. Finally, we repeat our calculations using different data sources. We compute the index using ϕ_i estimated from the CPS flows data (Figure B8). We also repeat our analysis at the 2-digit industry level using the HWOL ads data. Figure B9 paints a very similar picture to that obtained from JOLTS. Both $\mathcal{M}_t$ and $\mathcal{M}_t^x$ are somewhat higher for HWOL than for JOLTS, but peak and decline in a very similar fashion. The contribution of mismatch to the increase in unemployment rate is only slightly larger with HWOL—between 0.79 and 0.88 percentage points.

We also report additional results for occupational-level mismatch. First of all, Figure B11 presents indexes adjusted for one source of heterogeneity at a time for 2-digit occupations. Indexes for different values of α are plotted in Figure B12. Table B10 summarizes these results and shows that our baseline findings on occupation level mismatch of Section 5.2 are quite robust.

B.4 Adjustment in sectoral unemployment count

Let u_{it} be the number of unemployed workers at date t whose last job is in sector i, and U_{it} be the true number of unemployed actually searching in sector i at date t. Also let u_{it}^j be the number of unemployed whose last job is in sector i and who are searching in sector j. By definition, we have $u_{it} = \sum_{j=1}^{I} u_{it}^j$. The key unknown at each date t is the vector $\{U_{it}\}$.

From the panel dimension of CPS we observe h_{it}^j, the number of unemployed workers hired in sector j in period t whose last job was in sector i. Let the total number of hires in sector j in period t be h_t^j. Assume that the job-finding rate in sector j is the same for all unemployed, independent of the sector of provenance, with the sole exception if their previous job was in that same sector, in which case their job-finding rate is higher by a factor $\gamma_t \geq 1$, or:

$$\frac{h_{jt}^j}{u_{jt}^j} = (1 + \gamma_t)\,\frac{h_{it}^j}{u_{it}^j}, \text{ for } i \neq j. \tag{B4}$$

The average hiring rate of sector j is the total number of hires for j divided by the total number of unemployed looking in sector j or:

$$\frac{h_t^j}{U_{jt}} = \sum_{i \neq j} \left(\frac{h_{it}^j}{u_{it}^j} \right) \left(\frac{u_{it}^j}{U_{jt}} \right) + \left(\frac{h_{jt}^j}{u_{jt}^j} \right) \left(\frac{u_{jt}^j}{U_{jt}} \right).$$

Substituting $(B4)$ into the above equation delivers:

$$\frac{h_t^j}{U_{jt}} = \sum_{i \neq j} \left(\frac{h_{it}^j}{u_{it}^j} \right) \left(\frac{u_{it}^j}{U_{jt}} \right) + (1+\gamma_t) \frac{h_{it}^j}{u_{it}^j} \left(\frac{u_{jt}^j}{U_{jt}} \right).$$

Because the ratio h_{it}^j / u_{it}^j is the same across all $i \neq j$, we can pull it out of the sum above and obtain, after rearranging:

$$\frac{h_{it}^j}{u_{it}^j} = \begin{cases} \left(\frac{h_t^j}{U_{jt}} \right) \left[1 + \gamma_t \left(\frac{u_{jt}^j}{U_{jt}} \right) \right]^{-1} & \text{if } i \neq j \\ (1+\gamma_t) \left(\frac{h_t^j}{U_{jt}} \right) \left[1 + \gamma_t \left(\frac{u_{jt}^j}{U_{jt}} \right) \right]^{-1} & \text{if } i = j \end{cases} \tag{B5}$$

Since we do not observe u_{jt}^j / U_{jt}, we want to substitute it out. Note that

$$\frac{u_{jt}^j}{U_{jt}} = \frac{\frac{h_{jt}^j}{h_t^j} \left(\frac{1}{1+\gamma_t} \right)}{1 - \frac{h_{jt}^j}{h_t^j} \left(\frac{\gamma_t}{1+\gamma_t} \right)}$$

and using this expression in $(B5)$, we arrive at a relationship between the hiring rate from i to j and the average hiring rate in j:

$$\frac{h_{it}^j}{u_{it}^j} = \xi_{it}^j \cdot \frac{h_t^j}{U_{jt}} \tag{B6}$$

where

$$\xi_{it}^j = \begin{cases} 1 - \frac{h_{jt}^j}{h_t^j} \left(\frac{\gamma_t}{1+\gamma_t} \right) & \text{if } i \neq j \\ (1+\gamma_t) \left[1 - \frac{h_{jt}^j}{h_t^j} \left(\frac{\gamma_t}{1+\gamma_t} \right) \right] & \text{if } i = j \end{cases}$$

Rearranging equation $(B6)$ and summing across all j yields, at every t, the I equations:

$$u_{it} = \sum_{j=1}^{N} \frac{1}{\xi_{it}^j} \left(\frac{h_{it}^j}{h_t^j} \right) U_t^j$$

in the $(I+1)$ unknowns $\{U_{jt}\}, \gamma_t$. The last equation needed is the "aggregate consistency" condition

$$\sum_{j=1}^{I} U_{jt} = \sum_{j=1}^{I} u_{jt} \tag{B7}$$

stating that the true distribution of unemployed across sectors must sum to the observed total number of unemployed. We therefore have a system of $(I+1)$ equations in $(I+1)$ unknowns.

In our calculation of unemployment counts, to guarantee a non-negative solution to the linear system, we set to zero all entries in the transition matrices h_{it}^{j} which account for less than 5% of hires h_t^j in any given sector at any date t. We find that the estimated values of γ_t are all close to one.

Figures B16 and B17 plot the adjusted and unadjusted unemployment counts for some selected industries and occupations. As expected, for example, this correction reduces the number of unemployed workers searching in construction and increases that of those seeking jobs in healthcare.

B.5 Reweighting of HWOL vacancies

Let v_{irt}^{H} be the vacancies in the HWOL data for industry $i = 1, ..., I$ and region $r = 1, ..., R$ in month t. Let v_{irt}^{J} be the corresponding count for JOLTS vacancies. The objective is to reweigh monthly vacancies in HWOL to match those in JOLTS by industry and region (the only two common variables across data sets). We therefore solve, at every t, the following set of $(I \times R)$ equations

$$\begin{aligned} \sum_{i=1}^{I} v_{irt}^{H} \cdot \omega_{it} \cdot \omega_{rt} &= v_{rt}^{J} \\ \sum_{r=1}^{R} v_{irt}^{H} \cdot \omega_{it} \cdot \omega_{rt} &= v_{it}^{J} \end{aligned}$$

for the $(I \times R)$ vector of weights $\{\omega_{it}, \omega_{rt}\}$. Our solution algorithm imposes that weights must be positive, but this constraint is never binding in practice. Table B14 reports the average estimates of these weights over 2005-2006 and 2010-2011. We then compute reweighed vacancy counts by occupation o in month t as

$$v_{ot}^{H} = \sum_{i=1}^{I} \sum_{r=1}^{R} \omega_{it} \cdot \omega_{rt} \cdot v_{oirt}^{H}.$$

Our reweighed occupational mismatch index of Figure B18 is based on this revised vacancy count.

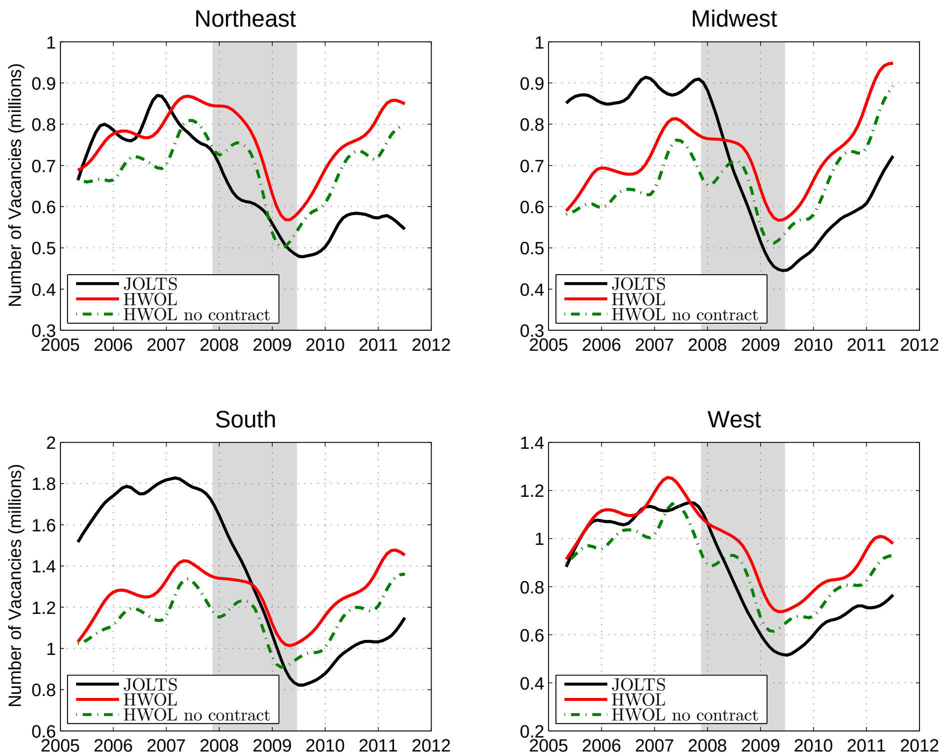
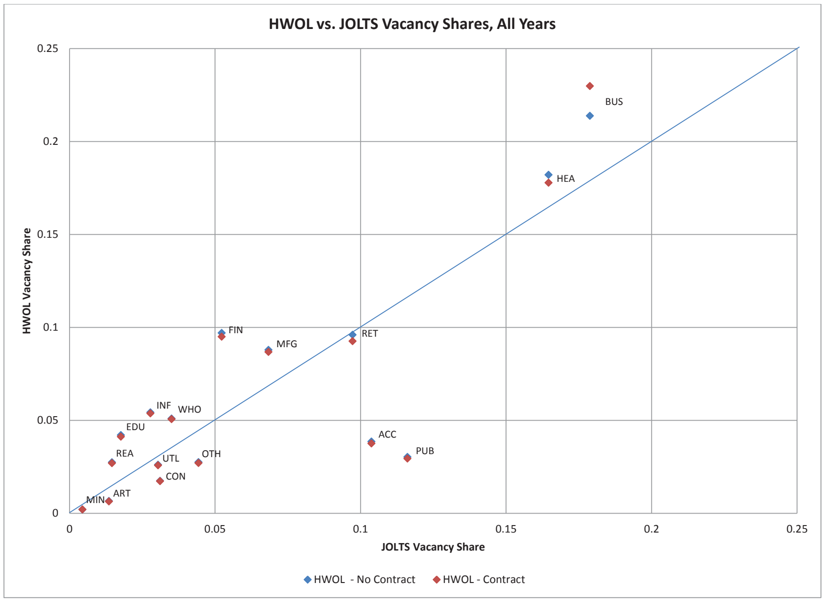


Figure B1: Comparison between the JOLTS and the HWOL (The Conference Board Help Wanted OnLine Data Series). Top-left panel: Northeast, Top-right panel: Midwest, Bottom-left panel: South, Bottom-right panel: West.



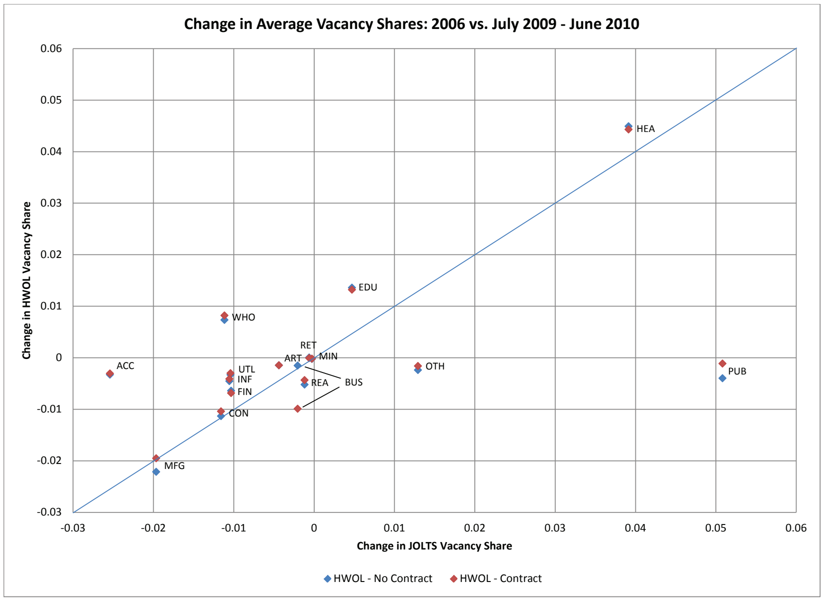


Figure B2: Top panel: comparison between vacancy shares in the JOLTS and HWOL (The Conference Board Help Wanted OnLine Data Series) for the May 2005 to June 2011 period. Bottom panel: change in average vacancy shares from 2006 to July 2009-June 2010 in the JOLTS and the HWOL. See Table B1 for an explanation of industry labels.

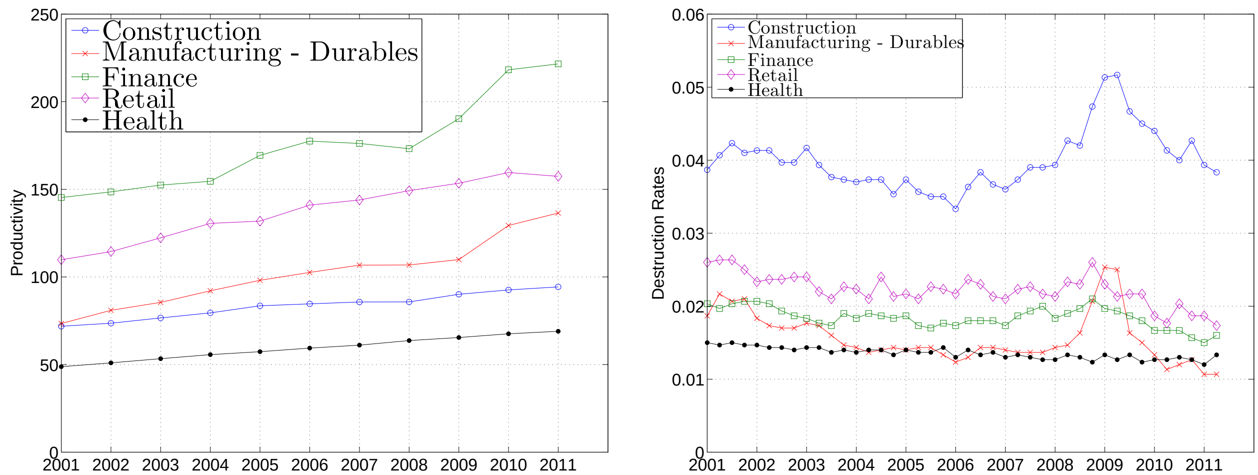


Figure B3: Productivity levels (left panel) and job destruction rates (right panel) for selected industries. Source: BEA and BLS for productivity levels and BED for job destruction rates.

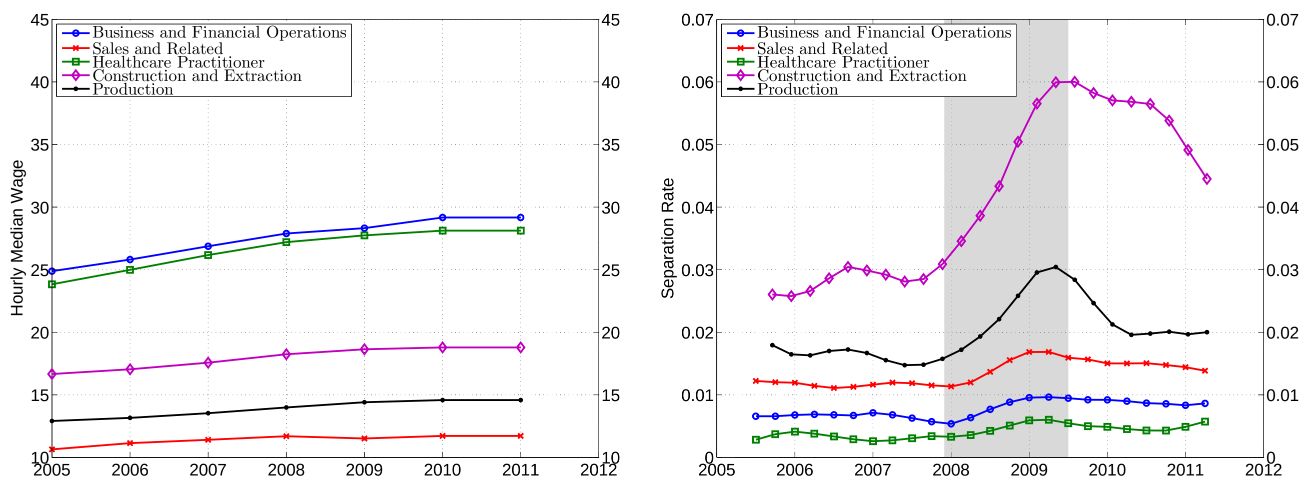


Figure B4: Wages (left panel) and job separation rates (right panel) for selected occupations. Source: OES for wages and CPS for job separation rates.

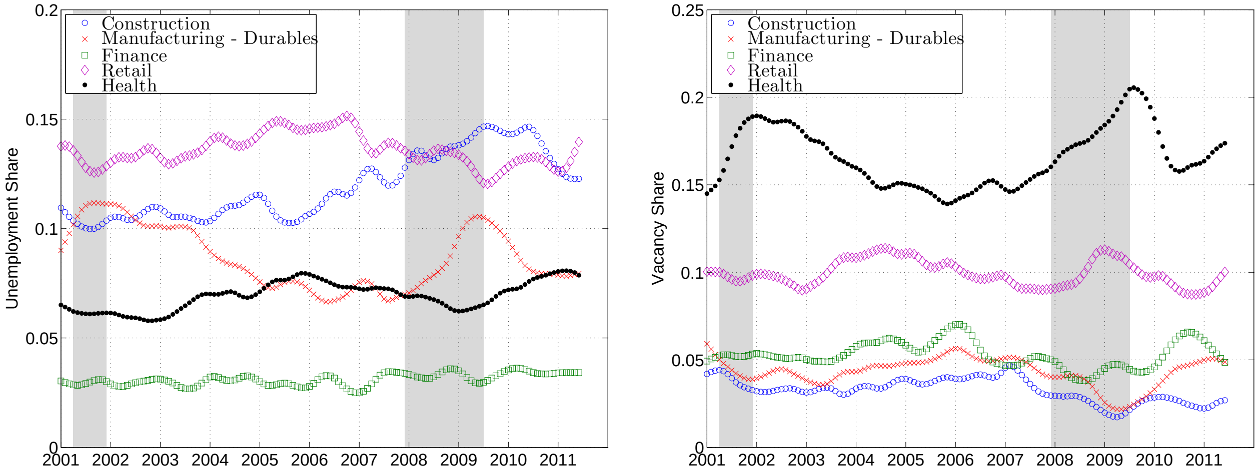


Figure B5: Unemployment and vacancy shares by selected industry.

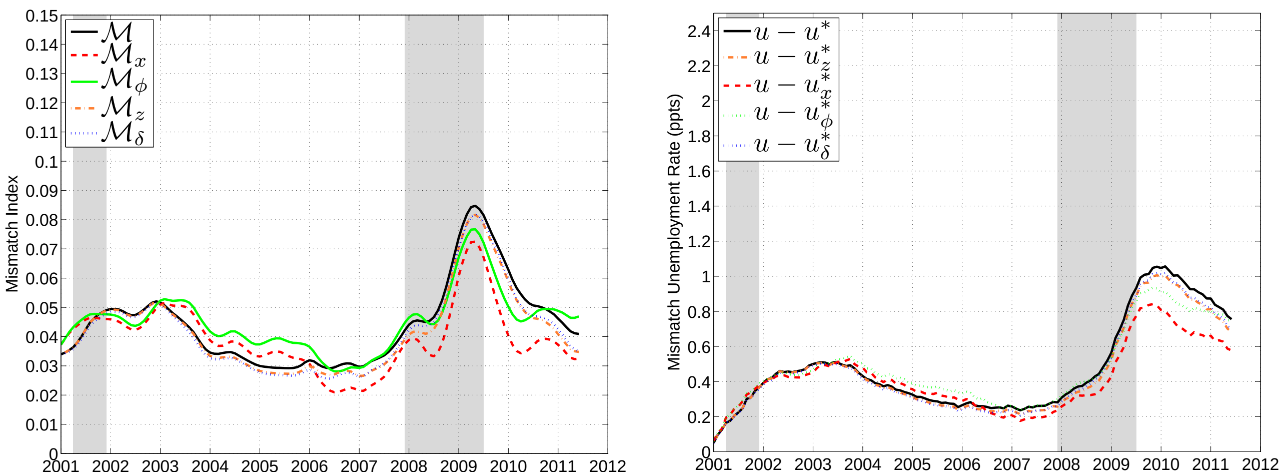


Figure B6: Mismatch indexes $\mathcal{M}_t$, $\mathcal{M}_{xt}$, $\mathcal{M}_{\phi t}$, $\mathcal{M}_{zt}$, and $\mathcal{M}_{\delta t}$ by industry (left panel) and the corresponding mismatch unemployment rates (right panel).

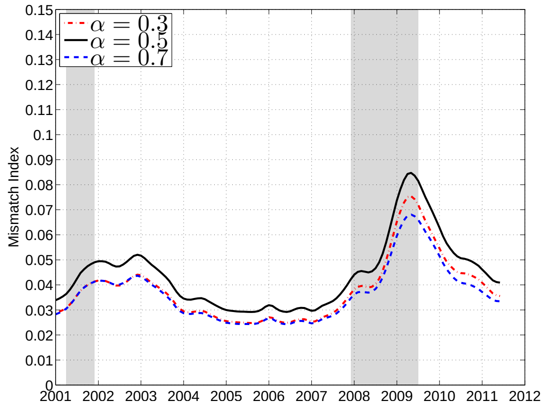


Figure B7: Mismatch index $\mathcal{M}_t$ by industry (left panel) and the corresponding mismatch unemployment rates (right panel) for various values of α.

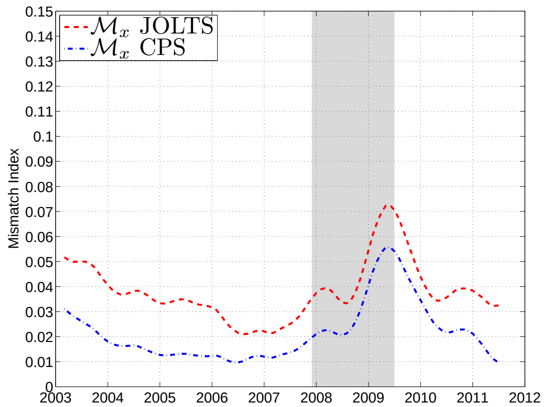


Figure B8: Mismatch index $\mathcal{M}_{xt}$ by industry (left panel) and the corresponding mismatch unemployment rates (right panel) using CPS measure of hires from unemployment.

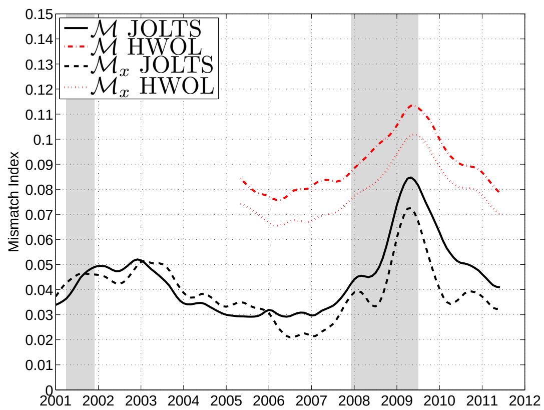


Figure B9: Mismatch indexes $\mathcal{M}_t$ (left panel) and the corresponding mismatch unemployment rates (right panel) using the JOLTS and HWOL (The Conference Board Help Wanted OnLine Data Series).

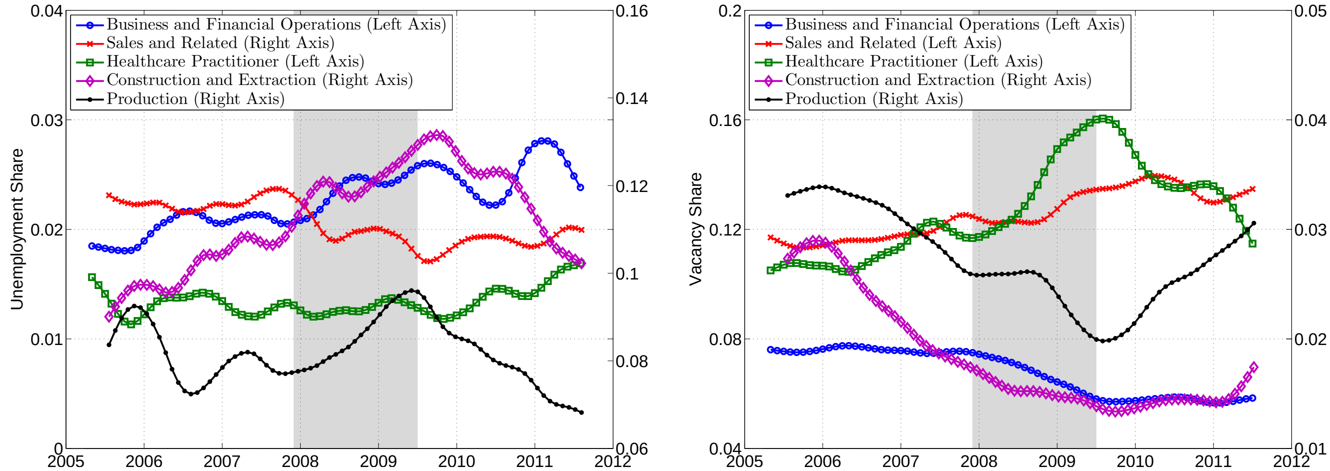


Figure B10: Unemployment and vacancy shares by selected occupation.

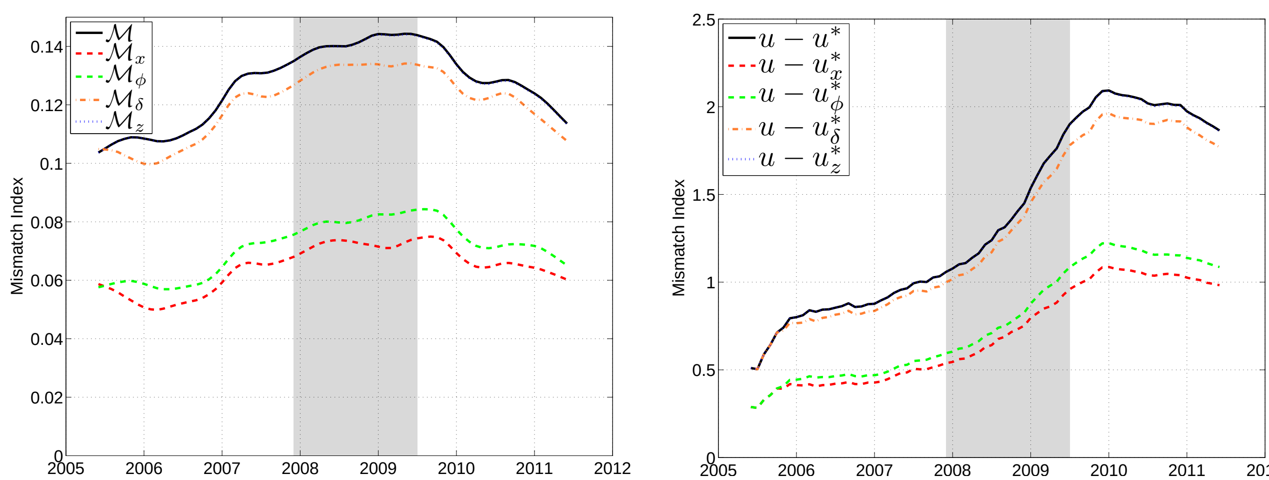


Figure B11: Mismatch indexes $\mathcal{M}_t$, $\mathcal{M}_{xt}$, $\mathcal{M}_{\phi t}$, $\mathcal{M}_{zt}$, and $\mathcal{M}_{\delta t}$ by occupation (left panel) and the corresponding mismatch unemployment rates (right panel).

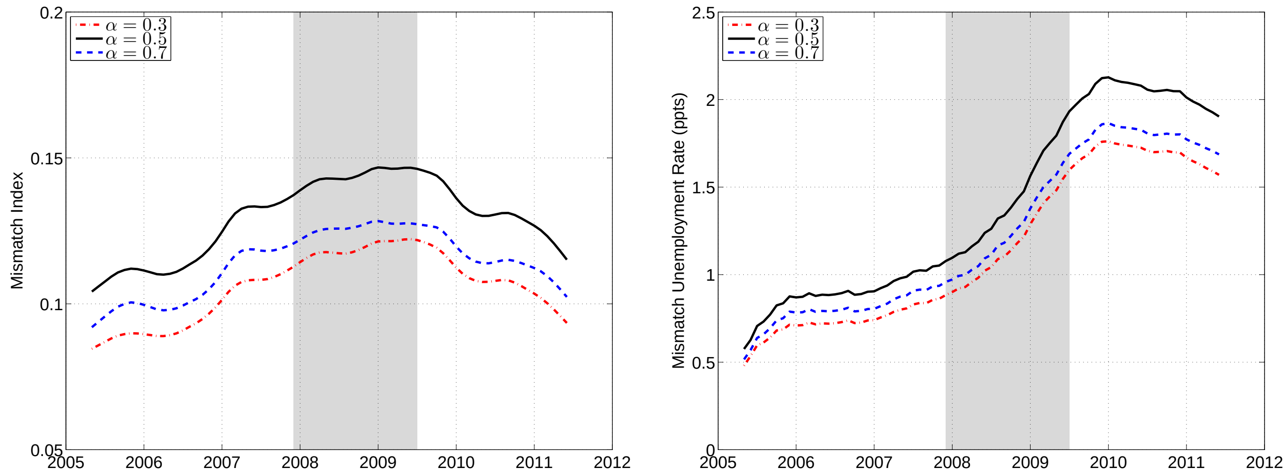


Figure B12: Mismatch index $\mathcal{M}_t$ by occupation (left panel) and the corresponding mismatch unemployment rates (right panel) for various values of α.

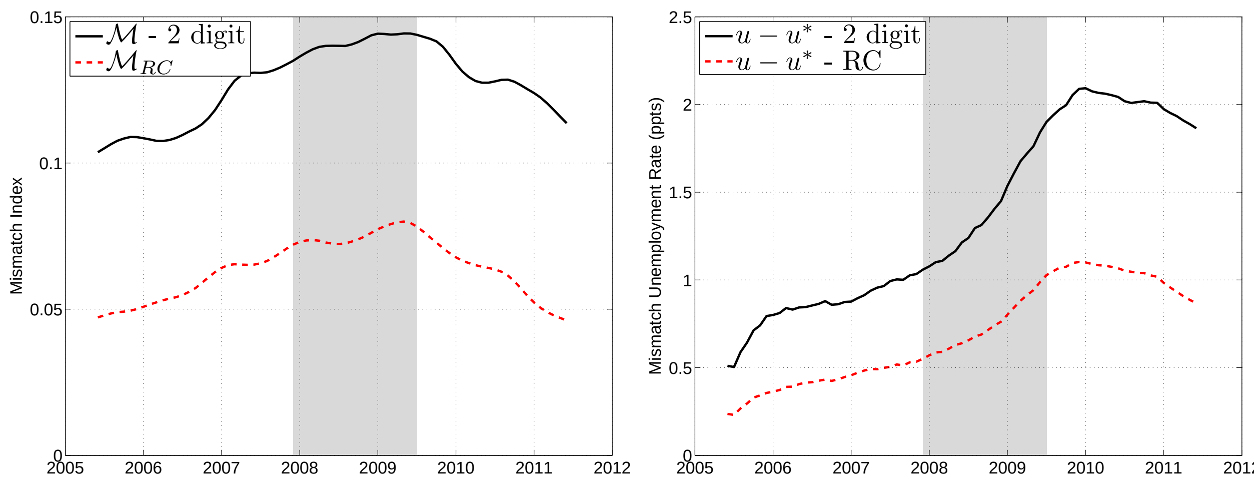


Figure B13: Mismatch indexes $\mathcal{M}$ across four occupations groups (routine/cognitive, manual/non-manual, and across 2-digit occupations (left panel). Corresponding mismatch unemployment rates (right panel).

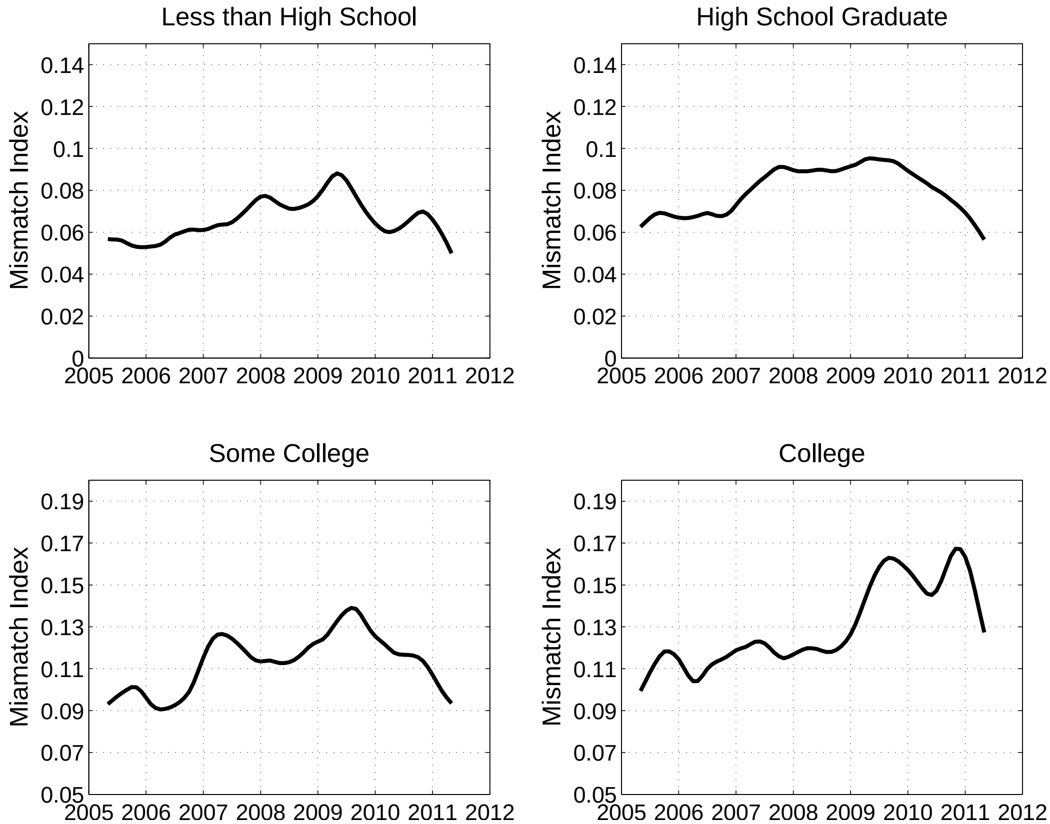


Figure B14: Mismatch indexes ($\mathcal{M}_t$) by occupation for different education groups.

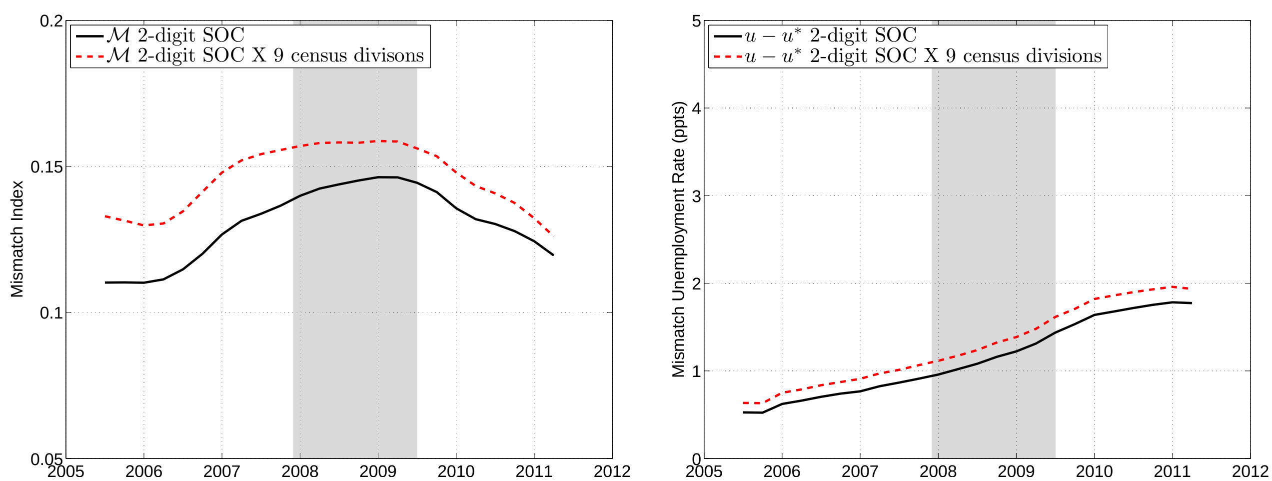


Figure B15: Mismatch index $\mathcal{M}_t$ by occupation and location (left panel) and the corresponding mismatch unemployment rates (right panel).

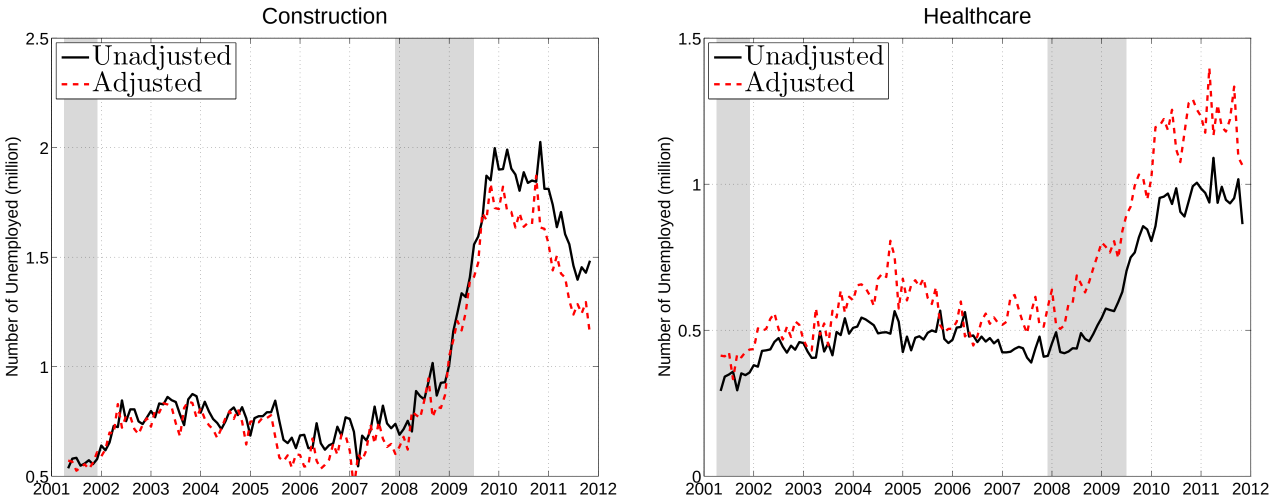


Figure B16: Adjusted unemployment counts for selected industries.

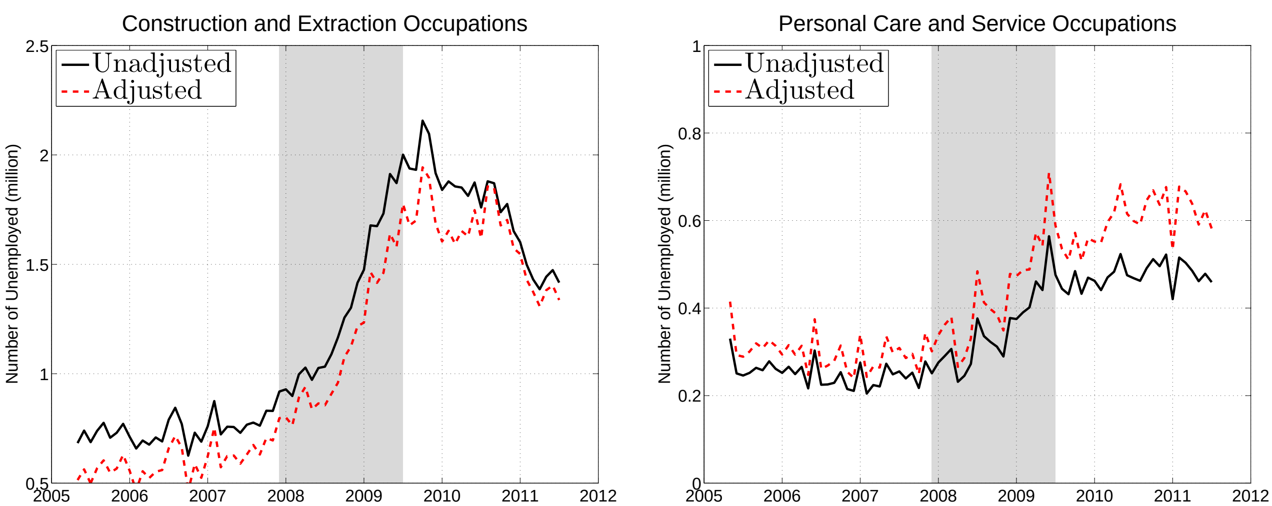


Figure B17: Adjusted unemployment counts for selected occupations.

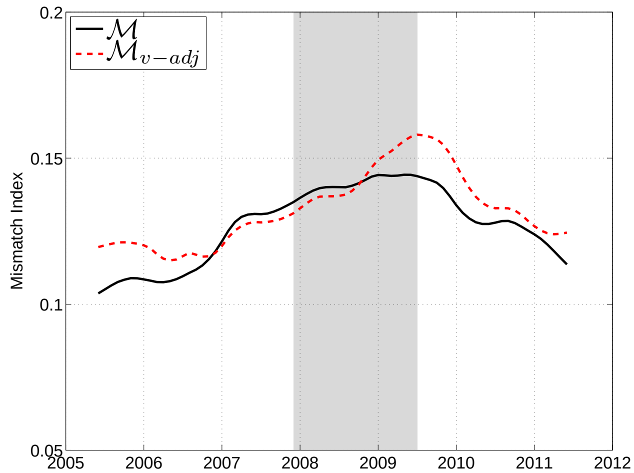


Figure B18: Mismatch index by 2-digit occupation: unadjusted index and index computed with reweighted HWOL vacancies

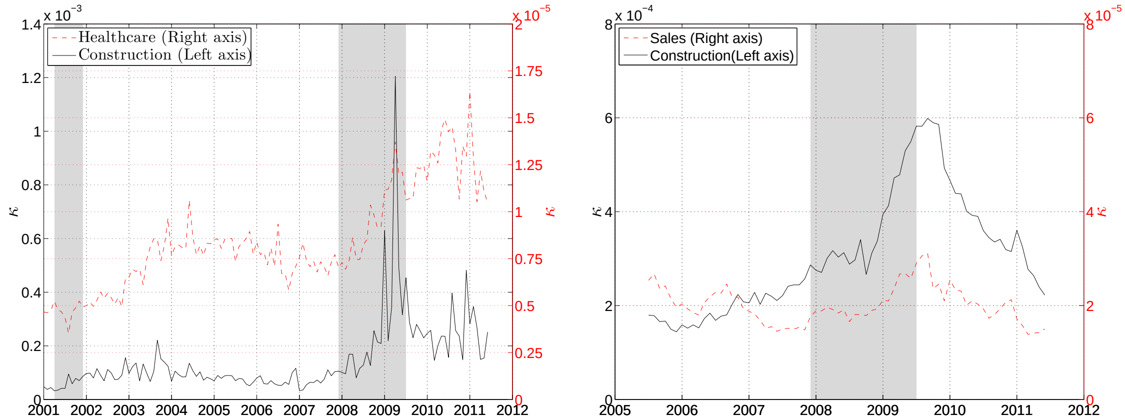


Figure B19: Time series of κ for two selected industries: construction and healthcare (left panel) and two selected occupations: construction and extraction occupations, and sales and related occupations (right panel). The cost is normalized by average annual labor productivity of the industry (annual wage for the occupation).

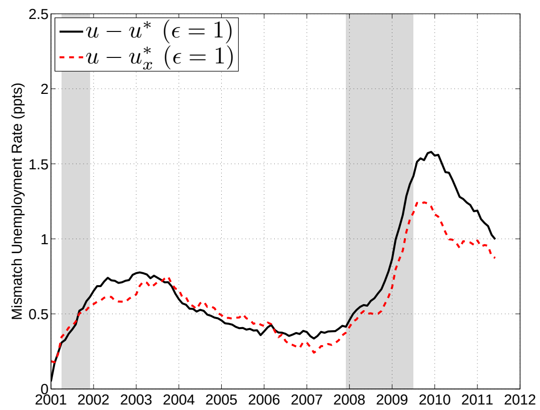


Figure B20: Mismatch unemployment with $\mathcal{M}_t^{v^*}$ at the industry level using endogenous vacancies specification with JOLTS.

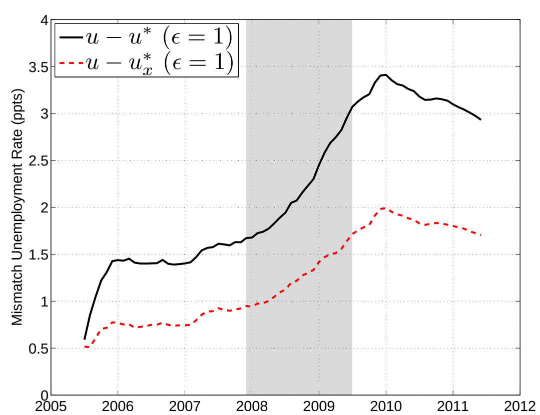


Figure B21: Mismatch unemployment with $\mathcal{M}_t^{v^*}$ at the occupation level using endogenous vacancies specification with the HWOL (The Conference Board Help Wanted OnLine Data Series).

Code	Industry
ACC	Accomodation and Food Services
ART	Arts, Entertainment and Recreation
CON	Construction
EDU	Education Services
FIN	Finance and Insurance
PUB	Government
HEA	Health Care and Social Assistance
INF	Information
MFG	Manufacturing-Durable Goods
MFG	Manufacturing-Nondurable Goods
MIN	Mining
OTH	Other Services
BUS	Professional and Business Services
REA	Real Estate and Rental and Leasing
RET	Retail Trade
UTL	Transportation, Warehousing and Utilities
WHO	Wholesale Trade

Table B1: Industry classification in the JOLTS. The codes in the left column are those used in Figure B2.

Code	Occupation	Classification
110000	Management Occupations	Cognitive/Non-routine
130000	Business and Financial Operations Occupations	Cognitive/Non-routine
150000	Computer and Mathematical Occupations	Cognitive/Non-routine
170000	Architecture and Engineering Occupations	Cognitive/Non-routine
190000	Life, Physical, and Social Science Occupations	Cognitive/Non-routine
210000	Community and Social Service Occupations	Cognitive/Non-routine
230000	Legal Occupations	Cognitive/Non-routine
250000	Education, Training, and Library Occupations	Cognitive/Non-routine
270000	Arts, Design, Entertainment, Sports, and Media Occupations	Cognitive/Non-routine
290000	Healthcare Practitioners and Technical Occupations	Cognitive/Non-routine
310000	Healthcare Support Occupations	Manual/Non-routine
330000	Protective Service Occupations	Manual/Non-routine
350000	Food Preparation and Serving Related Occupations	Manual/Non-routine
370000	Building and Grounds Cleaning and Maintenance Occupations	Manual/Non-routine
390000	Personal Care and Service Occupations	Manual/Non-routine
410000	Sales and Related Occupations	Cognitive/Routine
430000	Office and Administrative Support Occupations	Cognitive/Routine
470000	Construction and Extraction Occupations	Manual/Routine
490000	Installation, Maintenance, and Repair Occupations	Manual/Routine
510000	Production Occupations	Manual/Routine
530000	Transportation and Material Moving Occupations	Manual/Routine

Table B2: 2-digit SOC Codes used in our empirical analysis. The classification in the right column is that used in Figure B13.

Code	Occupation
111000	Top Executives
113000	Operations Specialties Managers
119000	Other Management Occupations
131000	Business Operations Specialists
132000	Financial Specialists
151000	Computer Occupations
211000	Counselors, Social Workers, and Other Community and Social Service Specialists
252000	Preschool, Primary, Secondary, and Special Education School Teachers
272000	Entertainers and Performers, Sports and Related Workers
291000	Health Diagnosing and Treating Practitioners
311000	Nursing, Psychiatric, and Home Health Aides
339000	Other Protective Service Workers
352000	Cooks and Food Preparation Workers
353000	Food and Beverage Serving Workers
359000	Other Food Preparation and Serving Related Workers
372000	Building Cleaning and Pest Control Workers
373000	Grounds Maintenance Workers
399000	Other Personal Care and Service Workers
411000	Supervisors of Sales Workers
412000	Retail Sales Workers
413000	Sales Representatives, Services
419000	Other Sales and Related Workers
433000	Financial Clerks
434000	Information and Record Clerks
435000	Material Recording, Scheduling, Dispatching, and Distributing Workers
436000	Secretaries and Administrative Assistants
439000	Other Office and Administrative Support Workers
452000	Agricultural Workers
472000	Construction Trades Workers
493000	Vehicle and Mobile Equipment Mechanics, Installers, and Repairers
499000	Other Installation, Maintenance, and Repair Occupations
512000	Assemblers and Fabricators
514000	Metal Workers and Plastic Workers
519000	Other Production Occupations
533000	Motor Vehicle Operators
537000	Material Moving Workers

Table B3: 3-digit SOC Codes used in our empirical analysis.

	Aggregate regressions				Panel regressions	
	JOLTS		HWOL		Industry (JOLTS)	Occupation (HWOL)
	OLS	GMM	OLS	GMM	OLS	OLS
JOLTS Hires	0.654	0.661	–	–	0.532	–
	(0.010)	(0.037)	–	–	(0.013)	–
Sample Size	126	126	–	–	2,142	–
CPS Hires	0.318	0.298	0.332	0.536	0.241	0.279
	(0.017)	(0.136)	(0.038)	(0.059)	(0.014)	(0.016)
Sample Size	126	126	72	72	404	370

Table B4: OLS and GMM estimates of the vacancy share α using the JOLTS and HWOL datasets. S.E. in parenthesis. See Section B.2 for details.

	JOLTS		HWOL	
	α	σ	α	σ
JOLTS Hires	0.576	0.152	-	-
	[0.542,0.603]	[0.051,0.242]		
CPS Hires	0.301	0.18	0.239	-0.108
	[0.267,0.350]	[0.08,0.303]	[0.194,0.291]	[-0.226,0.004]

Table B5: Estimates of the vacancy share α and CES substitutability parameter σ, using industry and occupation level data. 95-5 confidence intervals computed via bootstrap. Sample sizes are the same as in Table B4.

Industry	ϕ^{pre}	ϕ^{post}
Mining	1.71	1.36
Arts	1.69	1.87
Construction	1.66	1.73
Accommodations	1.53	1.60
Retail	1.47	1.46
Professional and Business Services	1.43	1.45
Real Estate	1.41	1.22
Wholesale	1.21	1.35
Other	1.14	1.16
Transportation and Utilities	1.14	1.16
Manufacturing - Nondurables	0.96	1.00
Education	0.94	1.02
Health	0.93	1.05
Government	0.87	0.89
Finance	0.85	0.73
Manufacturing - Durables	0.84	0.78
Information	0.76	0.70

Table B6: Estimates of industry-specific match efficiencies using hires from the JOLTS.

Industry Groups	Industry	ϕ^{pre}	ϕ^{post}
Group 1	Construction Mining	0.50	0.55
Group 2	Manufacturing Other Transportation and Utilities	0.42	0.44
Group 3	Accommodations Arts Professional and Business Services Retail Wholesale	0.38	0.39
Group 4	Education Finance Government Health Information Real Estate	0.33	0.33

Table B7: Estimates of industry-specific match efficiencies using hires from the CPS.

Occupation Groups	Occupation	ϕ^{pre}	ϕ^{post}
Service	Protective Service Occupations Food Preparation and Serving Related Occupations Building and Grounds Cleaning and Maintenance Occupations Personal Care and Service Occupations	0.58	0.63
Natural Resources, Construction and Maintenance	Construction and Extraction Occupations Installation, Maintenance, and Repair Occupations	0.56	0.63
Production, Transportation and Material Moving	Production Occupations Transportation and Material Moving Occupations	0.48	0.52
Sales and Office	Sales and Related Occupations Office and Administrative Support Occupations	0.37	0.35
Management, Professional and Related	Management Occupations Business and Financial Operations Occupations Computer and Mathematical Occupations Architecture and Engineering Occupations Life, Physical, and Social Science Occupations Community and Social Service Occupations Legal Occupations Education, Training, and Library Occupations Arts, Design, Entertainment, Sports, and Media Occupations Healthcare Practitioners and Technical Occupations Healthcare Support Occupations	0.32	0.33

Table B8: Estimates of occupation-specific match efficiencies using hires from the CPS.

	Index	$u_{06} - u^*_{06}$	$u_{10.09} - u^*_{10.09}$	$\Delta(u - u^*)$	$\Delta(u - u^*)/\Delta u$
JOLTS Hires	$\mathcal{M}$	0.26	1.01	0.75	13.9%
	$\mathcal{M}_x$	0.22	0.84	0.59	11.0%
	$\mathcal{M}_\phi$	0.29	0.92	0.63	11.7%
	$\mathcal{M}_z$	0.24	0.96	0.72	13.3%
	$\mathcal{M}_\delta$	0.23	0.98	0.74	13.7%
	$\mathcal{M}_{u-adj}$	0.25	0.89	0.65	11.9%
	$\mathcal{M}(\alpha = 0.3)$	0.22	0.89	0.67	12.4%
	$\mathcal{M}(\alpha = 0.5)$	0.26	1.01	0.75	13.9%
	$\mathcal{M}(\alpha = 0.7)$	0.22	0.82	0.60	11.1%
	$\mathcal{M}_x^{break}$	0.25	0.92	0.67	12.4%
	$\mathcal{M}^{v^*}(\varepsilon = 1.0)$	0.38	1.52	1.14	21.1%
	$\mathcal{M}_x^{v^*}(\varepsilon = 1.0)$	0.34	1.24	0.90	16.6%
CPS Hires	$\mathcal{M}$	0.27	1.03	0.77	12.4%
	$\mathcal{M}_x$	0.10	0.61	0.51	9.4%
HWOL	$\mathcal{M}$	0.63	1.51	0.88	16.3%
	$\mathcal{M}_x$	0.56	1.35	0.79	14.7%

Table B9: Changes in mismatch unemployment at the industry level. All the changes are calculated as the difference between October 2009 and the average of 2006. Note that $\Delta u = 5.4$ percentage points.

	Index	$u_{06} - u^*_{06}$	$u_{10.09} - u^*_{10.09}$	$\Delta(u - u^*)$	$\Delta(u - u^*)/\Delta u$
2-digit	$\mathcal{M}$	0.85	2.00	1.14	21.3%
	$\mathcal{M}_x$	0.42	1.02	0.60	11.1%
	$\mathcal{M}_\phi$	0.46	1.15	0.69	12.8%
	$\mathcal{M}_z$	0.85	2.00	1.15	21.2%
	$\mathcal{M}_\delta$	0.80	1.86	1.05	19.5%
	$\mathcal{M}_{u-adj}$	0.84	2.00	1.16	21.4%
	$\mathcal{M}_{v-adj}$	0.93	2.12	1.19	22.1%
	$\mathcal{M}(\alpha = 0.3)$	0.72	1.68	0.96	17.8%
	$\mathcal{M}(\alpha = 0.5)$	0.85	2.00	1.14	21.3%
	$\mathcal{M}(\alpha = 0.7)$	0.79	1.77	0.98	18.1%
	$\mathcal{M}_x^{break}$	0.42	0.98	0.56	10.4%
	$\mathcal{M}^{v^*}(\varepsilon = 1.0)$	1.41	3.20	1.79	33.1%
	$\mathcal{M}_x^{v^*}(\varepsilon = 0.5)$	1.08	2.60	1.52	28.1%
	$\mathcal{M}_x^{v^*}(\varepsilon = 1.0)$	0.74	1.81	1.07	19.7%
3-digit	$\mathcal{M}$	1.33	2.91	1.58	29.3%
	$\mathcal{M}_x$	0.79	1.73	0.94	17.4%
	$\mathcal{M}_\phi$	0.83	1.85	1.02	18.8%
	$\mathcal{M}_z$	1.33	2.91	1.58	22.2%
	$\mathcal{M}_\delta$	1.29	2.80	1.50	27.8%

Table B10: Changes in mismatch unemployment at the occupation level. All the changes are calculated as the difference between October 2009 and the average of 2006. Note that $\Delta u = 5.4$ percentage points.

Index	$u_{Q1.01} - u^*_{Q1.01}$	$u_{06.03} - u^*_{06.03}$	$\Delta(u - u^*)$	$\Delta(u - u^*)/\Delta u$
$\mathcal{M}$	0.09	0.50	0.41	22.8%
$\mathcal{M}_x$	0.10	0.50	0.41	21.7%
$\mathcal{M}_{u-adj}$	0.11	0.43	0.32	17.8%
$\mathcal{M}_x^{v^*}(\varepsilon = 1.0)$	0.20	0.70	0.50	26.8%

Table B11: Changes in mismatch unemployment at the industry level for the 2001 recession. All the changes are calculated as the difference between June 2003 (month in which the unemployment rate peaked for the 2001 recession) and the average of 2001Q1. Note that $\Delta u = 1.8$ percentage points.

	July 2005-June 2007		July 2007-June 2009		July 2009-June 2011	
INDUSTRY	UN	UD	UN	UD	UN	UD
Agriculture and Mining	0.17	0.08	0.17	0.07	0.15	0.06
Construction	0.16	0.08	0.15	0.08	0.13	0.07
Manufacturing	0.21	0.10	0.18	0.08	0.16	0.08
Trade	0.24	0.10	0.23	0.10	0.21	0.10
Transportation and Utilities	0.23	0.10	0.19	0.09	0.17	0.08
Information	0.21	0.10	0.18	0.08	0.17	0.09
Financial	0.21	0.09	0.19	0.09	0.17	0.09
Professional Business Services	0.23	0.10	0.20	0.09	0.18	0.08
Education and Health	0.25	0.11	0.24	0.10	0.21	0.10
Leisure	0.29	0.11	0.25	0.10	0.24	0.10
Other	0.28	0.11	0.25	0.11	0.20	0.10
Public Administration	0.25	0.11	0.25	0.09	0.20	0.10
All	0.23	0.10	0.20	0.09	0.18	0.08

Table B12: UN and UD monthly flow rates (fraction of the unemployment pool) by industry.

	July 2005-June 2007		July 2007-June 2009		July 2009-June 2011	
OCCUPATION	UN	UD	UN	UD	UN	UD
Management	0.15	0.07	0.16	0.07	0.14	0.07
Business and Financial Operations	0.19	0.11	0.17	0.08	0.14	0.06
Computer and Mathematical	0.19	0.11	0.17	0.08	0.14	0.08
Architecture and Engineering	0.20	0.10	0.17	0.09	0.13	0.06
Life, Physical and Social Science	0.20	0.12	0.17	0.10	0.17	0.07
Community and Social Service	0.29	0.13	0.19	0.08	0.18	0.10
Legal	0.20	0.06	0.18	0.07	0.18	0.08
Education, Training, and Library	0.25	0.10	0.21	0.09	0.20	0.09
Arts, Design, Entertainment, Sports, and Media	0.20	0.08	0.19	0.08	0.18	0.08
Healthcare Practitioners and Technical	0.21	0.09	0.23	0.10	0.18	0.08
Healthcare Support	0.29	0.12	0.27	0.12	0.22	0.09
Protective Service	0.25	0.11	0.23	0.07	0.21	0.09
Food Preparation and Serving Related	0.29	0.11	0.25	0.10	0.24	0.09
Building and Grounds Cleaning and Maintenance	0.28	0.11	0.25	0.10	0.23	0.10
Personal Care and Service	0.30	0.13	0.26	0.11	0.26	0.10
Sales and Related	0.26	0.10	0.25	0.10	0.22	0.09
Office and Administrative Support	0.24	0.10	0.22	0.09	0.20	0.08
Construction and Extraction	0.17	0.08	0.16	0.08	0.13	0.07
Installation, Maintenance, and Repair	0.21	0.10	0.16	0.08	0.14	0.07
Production	0.22	0.10	0.18	0.08	0.16	0.08
Transportation and Material Moving	0.23	0.10	0.19	0.09	0.19	0.09
All	0.23	0.10	0.21	0.09	0.18	0.08

Table B13: UN and UD monthly flow rates (fraction of the unemployment pool) by occupation.

		Weight 2005-2006	Weight 2010-2011
Industry			
	Accomodation and Food Services	2.25	2.43
	Arts, Entertainment and Recreation	1.07	1.03
	Construction	1.42	1.32
	Education Services	0.44	0.55
	Finance and Insurance	0.49	0.56
	Government	2.94	2.35
	Health Care and Social Assistance	0.79	0.83
	Information	0.49	0.58
	Manufacturing-Durable Goods	0.81	0.64
	Manufacturing-Nondurable Goods	0.75	0.63
	Mining	0.82	1.23
	Other Services	1.34	1.14
	Professional and Business Services	0.34	0.35
	Real Estate and Rental and Leasing	0.56	0.52
	Retail Trade	0.92	1.04
	Transportation, Warehousing and Utilities	1.00	1.07
	Wholesale Trade	0.61	0.73
Region			
	Northeast	0.90	0.99
	West	1.18	0.97
	Southwest	0.68	0.92
	South	1.17	1.23

Table B14: Estimated weights which equalize monthly JOLTS and HWOL (The Conference Board Help Wanted OnLine Data Series) vacancy counts by industry and region (average weight is normalized to one each month).