Measuring Mismatch in the U.S. Labor Market*

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: OCTOBER 2011

Abstract

This paper measures mismatch between job-seekers and vacancies in the U.S. labor market. Mismatch is defined as the distance between the observed allocation of unemployed workers across sectors and the optimal allocation that solves a planner's problem. The planner's allocation rule requires (productive and matching) efficiency-weighted vacancy-unemployment ratios to be equated across sectors. More severe mismatch between vacant jobs and idle workers translates into higher unemployment by reducing the aggregate job-finding rate. In our empirical analysis, we use two sources of cross-sectional data on vacancies, JOLTS and HWOL, together with unemployment data from the CPS. We find that mismatch across industries and occupations accounts for 0.6 to 1.7 percentage points of the recent rise (by about five percentage points) in the U.S. unemployment rate, whereas geographical mismatch plays no role. The share of the rise in unemployment explained by mismatch is increasing in the education level.

*We wish to thank Michele Boldrin, Björn Brügemann, Mark Gertler, John Haltiwanger, Marianna Kudlyak, Rob Shimer, and numerous seminar participants for helpful comments. Grant Graziani, Victoria Gregory, Scott Nelson and Christina Patterson provided excellent research assistance. 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.

1 Introduction

The unemployment rate in the U.S. rose from 4.7% in December 2007 to 10.1% in October 2009, and has subsequently stayed above 9.0% through most of 2010 and 2011. This high unemployment has sparked a vibrant debate among economists and policymakers. The main point of contention is the nature of this persistent rise and, therefore, the appropriate policy response, if any.

A deeper look into flows into and out of unemployment shows that, while the rate of job loss has now returned to its pre-recession level, the job-finding rate is still significantly below its 2006 level. Any credible theory accounting for the recent dynamics in unemployment must therefore operate through a persistently lower exit rate from unemployment. One such theory is that unemployment is still high because of 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 would lower the aggregate job-finding rate.

The mismatch hypothesis seems, at a first pass, plausible because it is potentially coherent with three features of the Great Recession. First, a substantial fraction of job losses in this downturn was concentrated in construction, manufacturing, real estate, and finance, whereas vacancies –while decreasing across the board– refrained from dropping sharply only in a handful of industries (most notably, health care). Second, the depressed housing market may have slowed down geographical mobility of labor: if homeowners expect house prices to recover, they may delay the sale of their house –a necessary condition for mobility. Third, 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 given level of vacancies, the current level of aggregate unemployment is higher than that implied by the historical relationship between vacancies and unemployment.[1] Lack of coincidence between unemployment and vacancies across labor markets is consistent with this shift.

In this paper, we develop a theoretical framework to conceptualize the notion of mismatch unemployment. We then use this framework, together with disaggregated data on the distribution of vacancies and unemployed workers across occupations, industries, education levels and U.S. states, to measure how much of the recent rise in unemployment is due to mismatch.

To formalize the notion of mismatch, it is useful to envision the economy as comprising a large number of distinct labor markets, or sectors (e.g., segmented by industry, occupation, skill or education, geography, or a combination of these attributes). Each labor market is frictional, i.e., the hiring process within a labor market is governed by a matching function. To assess the existence of mismatch in the data we ask whether, given the distribution of vacancies observed in the economy, unemployed

[1] See, for example, Davis, Faberman, and Haltiwanger (2010), Elsby, Hobijn, and Şahin (2010), Hall (2010), and Daly, Hobijn, Şahin, and Valletta (2011).

workers are "misallocated". 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 can freely move unemployed workers across sectors*. The planner's cross-sectoral unemployment allocation rule dictates that (productive and matching) efficiency-weighted vacancy-unemployment ratios be equated across sectors.

We then construct a mismatch index that measures how many additional hires the ideal distribution of unemployed workers across sectors would generate relative to the observed equilibrium distribution. Through this index, we can define a counterfactual aggregate job-finding rate and quantify how much lower the unemployment rate would be in the absence of mismatch. The difference between the observed unemployment rate and the counterfactual unemployment rate based on the planner's allocation rule provides an estimate of *mismatch unemployment*. This formalization of mismatch unemployment follows, in essence, the same insight of the large literature on misallocation and productivity (Lagos, 2006; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Jones, 2011; Moll, 2011) and the literature on wedges (Chari, Kehoe, McGrattan, 2007).

Our strategy is to compare the empirical distribution of unemployment (the equilibrium outcome) to the allocation rule arising from the problem of a planner who has the ability to move labor across sectors at no cost. This approach allows to measure mismatch and its contribution to the recent rise of unemployment. We do not provide a structural model that incorporates all the potential sources of mismatch and delivers mismatch as an equilibrium outcome. Factors explaining the difference between the empirical allocation and the planner's allocation may include moving or retraining costs that an unemployed worker may incur when she searches in a different sector than her original one, as well as any other distortions originating, for instance, from incomplete insurance, imperfect information, wage rigidities, or various government policies. While we are not in the best position to identify its causes, we argue that studying mismatch for different definitions of sector (occupation, industry, education, geography) is informative about its roots.

We apply our analysis to the U.S. labor market and construct measures of mismatch across industries, occupations, education levels and geographic areas using vacancy data from the Job Openings and Labor Turnover Survey (JOLTS) and from the Conference Board's Help Wanted OnLine (HWOL) database, and unemployment data from the Current Population Survey (CPS).[2] We find that mismatch at the industry and occupation level increased during the recession and started to come down in 2010; an indication of a cyclical pattern for mismatch. Our calculations show that mismatch accounted for at most 0.6 to 1.7 percentage points of the total increase (around five percentage points) in the unemployment rate from the start of the recession to the unemployment peak in late 2009. We also calculate geographic mismatch measures and find no role for geographic mismatch in explaining the increase in the unemployment rate. This finding is consistent with other recent work that investigates

[2]In Şahin, Song, Topa and Violante (2011), we also apply our methodology to the U.K labor market.

the house-lock mechanism using different methods.[3] When we perform our study of occupational mismatch separately for different education groups, we find that the portion of the rise in unemployment explained by mismatch increases steeply with education. This result is consistent with the view that the human capital of the more highly educated is more specialized.

Our paper relates to an old, mostly empirical, literature that popularized the idea of mismatch (or what used to be called 'structural') unemployment in the 1980's when economists were struggling to understand why unemployment kept rising steadily in many European countries. 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. Padoa-Schioppa (1991) contains a number of empirical studies on mismatch and concludes that it was not an important explanation of the dynamics of European unemployment in the 1980s.[4] Within that literature, the closest paper to ours is Jackman and Roper (1987): they show, in a simple static model, that the optimal allocation of unemployment equates market tightness across sectors, and deviations from such allocations represent a measure of what they label "structural unemployment".

More recently, Barnichon and Figura (2011) have contributed to reviving this literature by showing that the variance of labor market tightness across sectors, suggestive of mismatch between unemployment and vacancies, can be analytically related to aggregate matching efficiency and, hence, can be a source of variation in the job-finding rate. Our approach is different and our scope broader, but we also show that changes in mismatch act as shifts in the aggregate matching function.

At a more theoretical level, Shimer (2007a) and Mortensen (2009) were the first to develop the idea that an economy with many separate labor markets, and misallocation of job-seekers and vacancies across markets, could be empirically consistent with the aggregate Beveridge curve. In this set-up, workers are randomly assigned to markets. Alvarez and Shimer (2010), Birchenall (2010), Carrillo-Tudela and Visscher (2010), and Hertz and Van Rens (2011) have all proposed dynamic models with explicit mobility decisions across labor markets where unemployed workers, in equilibrium, may be mismatched. While less amenable to measurement than our framework, these equilibrium models may be better suited to study the deeper causes of mismatch.

The rest 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 counterfactuals. Section 4 describes the data. Section 5 performs the empirical analysis. Section 6 concludes.

[3] See, for example, Molloy, Smith, and Wozniak (2010) and Schulhofer-Wohl (2010).

[4] Since then, it has become clear that explanations of European unemployment based on the interaction between technological changes in the environment and rigid labor market policies are more successful quantitatively (e.g., Ljungqvist and Sargent, 1998; Mortensen and Pissarides, 1999; Hornstein, Krusell and Violante, 2007).

2 Theoretical framework

We begin with a baseline economic environment where few simplifying assumptions lead to a straightforward derivation of the planner's optimal allocation rule of unemployed workers across sectors — the crucial building block of our empirical analysis. In Section 2.3 we describe a number of generalizations of the baseline model. In particular, we endogenize separations and labor force participation, and we allow for heterogeneous productivities and heterogeneous job destruction rates across sectors. In all these richer environments, the baseline planner's allocation rule generalizes in very intuitive ways.

2.1 Baseline 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. 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, $\sum_{i=1}^{I}(e_i + u_i) = 1$. It is useful to note explicitly that on-the-job search is ruled out, and that 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 $\Phi \cdot \phi_i \cdot m(u_i, v_i)$, 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.

Existing matches in sector i produce Z units of output, where Z is common across sectors. New matches produce only a fraction $\gamma < 1$ of output compared to existing matches –a stylized way to capture training costs for hiring unemployed workers. Matches are destroyed exogenously at rate δ, 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 $\mathbf{e} = \{e_1, ...e_I\}$ across markets (and hence the total number of unemployed workers u) is also given. Next, the planner chooses the number of unemployed workers to allocate in each labor market i. 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 is destroyed exogenously in each market i, determining next period's employment distribution $\{e_i'\}$ and stock of unemployed workers u'.

2.2 Planner's solution

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

$$\begin{aligned} V(\mathbf{e}; \mathbf{v}, \phi, Z, \delta, \Phi) &= \max_{\{u_i \geq 0\}} \sum_{i=1}^{I} Z(e_i + \gamma h_i) + \beta \mathbb{E}\left[V(\mathbf{e}'; \mathbf{v}', \phi', Z', \delta', \Phi')\right] \\ s.t. &: \\ \sum_{i=1}^{I} (e_i + u_i) &= 1 \qquad (1) \\ h_i &= \Phi\phi_i m(u_i, v_i) \qquad (2) \\ e_i' &= (1-\delta)(e_i + h_i) \qquad (3) \\ & \Gamma_{Z,\delta,\Phi}(Z', \delta', \Phi'; Z, \delta, \Phi), \ \Gamma_{\mathbf{v}}(\mathbf{v}'; \mathbf{v}, Z', \delta', \Phi'), \Gamma_{\phi}(\phi'; \phi) \qquad (4) \end{aligned}$$

The per period output for the planner is equal to $Z(e_i + \gamma h_i)$ in each market i. The first constraint (1) states that the planner has $1 - \sum_{i=1}^{I} e_i$ unemployed workers available to allocate across sectors.[5] Equation (2) 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 (3) describes separations and the determination of next period's distribution of active matches $\{e_i'\}$ in all sectors. Line (4) 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. The choice of how many unemployed workers u_i to allocate in market i yields the first-order condition

$$Z\gamma\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \beta\mathbb{E}\left[V_{e_i}'(\mathbf{e}'; \mathbf{v}', \phi', Z', \delta', \Phi')\right](1-\delta)\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu, \qquad (5)$$

where μ is the multiplier on constraint (1). 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 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 CRS.

[5] We imposed equality in this constraint because search is costless for the planner.

The Envelope condition with respect to the state e_i yields:

$$V_{e_i}(\mathbf{e};\mathbf{v},\phi,Z,\delta,\Phi) = Z - \mu + \beta(1-\delta)\mathbb{E}\left[V_{e_i}(\mathbf{e}';\mathbf{v}',\phi',Z',\delta',\Phi')\right], \tag{6}$$

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.[6] Using this result into (5), the optimal rule for the allocation of unemployed workers across sectors can be written as

$$\phi_1 m_{u_1}\left(\frac{v_1}{u_1^*}\right) = ... = \phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) = ... = \phi_I m_{u_I}\left(\frac{v_I}{u_I^*}\right), \tag{7}$$

where we have used the "*" to denote the optimal allocation. This is our key optimality condition for the allocation of unemployed workers across labor markets. It states that the higher vacancies and matching efficiency in market i, the more unemployed workers the planner wants searching in that market.

2.3 Generalizations

We develop three generalizations of our baseline model that will be useful in guiding the empirical analysis of Section 5. First, we allow productivities to differ across sectors. Here, we discuss two cases. One where sector-specific shocks are uncorrelated across sectors and independent of the aggregate shock; another where sectoral fluctuations in productivity are driven by the aggregate shock, but different sectors have different elasticities to this common factor. Second, we allow for exogenous match destruction rates to differ across sectors. Third, we let the planner choose whether to endogenously dissolve some existing matches and show that, under some conditions, it never chooses to do so. Throughout these extensions, we also allow the planner to choose the size of the labor force. We normalize to zero utility from non participation, and allow for disutility of search. All the derivations are contained in Appendix A1.

2.3.1 Heterogeneous productivities

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. Let the conditional distribution of the vector $\mathbf{z} = \{z_i\}$ be $\Gamma_z(\mathbf{z}',\mathbf{z})$ with a linear conditional mean function. The pdf of new vacancies across sectors $\{v_i\}$ is allowed to be correlated with the distribution of productivities $\{z_i\}$. The planner's allocation rule of unemployed workers across labor markets satisfies

$$z_1\phi_1 m_{u_1}\left(\frac{v_1}{u_1^*}\right) = ... = z_i\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) = ... = z_I\phi_I m_{u_I}\left(\frac{v_I}{u_I^*}\right), \tag{8}$$

a condition stating that the higher vacancies and matching and productive efficiency in market i, the more unemployed workers the planner wants searching in that market.

[6] We are also imposing the transversality condition $\lim_{t\to\infty}\beta^t(1-\delta)^t\mathbb{E}\left[V'_{e_i,t}\right] = 0$.

2.3.2 Heterogeneous sensitivities to the aggregate shock

In a classic paper disputing Lilien's (1982) sectoral-shift theory of unemployment, Abraham and Katz (1986) argue that, empirically, sectoral employment movements appear to be driven by aggregate shocks with different sectors having different sensitivities to the aggregate cycle. Here we show how the planner's allocation rule changes under this alternative interpretation of what drives sectoral labor demand shifts.

Let productivity in sector i be Z^{η_i} where η_i is a sector specific parameter measuring the elasticity to the aggregate shock Z. Let $\log Z$ follow a unit root process with innovation ε distributed as a $N\left(-\sigma_\varepsilon/2, \sigma_\varepsilon\right)$. To simplify the exposition, set $\gamma = 1$ and assume that δ is constant over time. The planner will allocate unemployed workers so to equalize

$$\frac{Z^{\eta_i}}{1-\beta\left(1-\delta\right)\exp\left(\eta_i\left(\eta_i-1\right)\frac{\sigma_\varepsilon}{2}\right)}\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) \tag{9}$$

across sectors. The new term in the denominator captures that the drift in future productivity in sector i depends on the variance of the aggregate shock proportionately to η_i because of the log-normality assumption. In essence, the effective rate at which the planner discounts future output becomes sector specific. With estimates of the elasticities $\{\eta_i\}$ and of the parameters of the stochastic process for Z in hand, the expression above can be easily computed.

Understanding the nature of sectoral fluctuations goes beyond the scope of this paper, and we do not make any attempt to contribute to this literature. The main lesson of this generalization is that our approach is valid under alternative views of what drives sectoral fluctuations. Different views simply lead to different measurements of the sectoral component of productivity in the planner's allocation rule.

2.3.3 Heterogeneous destruction rates

We now relax the assumption that the destruction rate δ is common across sectors. Consider the environment of Section 2.3.1. Denote the idiosyncratic component of the exogenous destruction rate in sector i as δ_i. Then, the survival probability of a match is $\left(1-\delta\right)\left(1-\delta_i\right)$. To simplify the exposition, set $\gamma = 1$, and assume that $\{Z, \delta, z_i, \delta_i\}$ all follow independent unit root processes.[7] In Appendix A1 we prove that the planner allocates idle labor to equalize

$$\frac{z_i}{1-\beta\left(1-\delta\right)\left(1-\delta_i\right)}\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right) \tag{10}$$

across sectors. The new term captures the fact that the expected output of an unemployed in sector i is discounted differently by the planner in different sectors because of the heterogeneity in the expected duration of matches.

[7] We can allow the pdf of new vacancies in sector i to be correlated with the destruction rate in that sector.

2.3.4 Endogenous separations

Consider the environment of Section 2.3.1 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. In Appendix A1 we show that, if the planner has always enough individuals to pull into (out of) unemployment from (into) out of the labor force, it will never choose to separate workers who are matched and producing, and the planner's allocation rule remains exactly as in equation (8).

2.4 Comparison between actual and optimal allocation: what do we measure?

Our approach to quantify the mismatch component of unemployment at date t is based on comparing the actual (equilibrium) distribution $\{u_{it}\}$ observed directly from the data to the optimal (planner's) distribution $\{u_{it}^*\}$ implied by (7) and its generalizations, for an (exogenously given) distribution of vacancies $\{v_{it}\}$ across sectors of the economy. This approach, where data are compared to an "ideal" allocation is at the heart of the growing literature on misallocation (e.g., Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Jones, 2011; Moll, 2011).

In equilibrium, there are a number of sources of misallocation that may induce $\{u_{it}\}$ to deviate from $\{u_{it}^*\}$ including imperfect information, wage rigidities, government policies, and moving/retraining costs. Under imperfect information, workers may be reluctant to move because they do not know where the vacancies are or what their prospects might be in the new location, occupation or industry. In the presence of wage rigidities, workers may choose not to move because wages deviate from productivity remaining relatively high (low) in the declining (expanding) sectors. An array of government interventions (e.g., generous unemployment benefits, housing and mortgage related policies, sector-specific taxes/transfers) may hamper mobility and be a source of misallocation. Moving or retraining costs associated to working in a new location, industry or occupation can obviously reduce mobility.

By following our approach, one does not need to model explicitly any of the sources of misallocation since the distribution $\{u_{it}\}$ comes straight from the data and the distribution $\{u_{it}^*\}$ is obtained from the problem of a planner who can freely move labor across sectors. The crucial advantage is that optimality can be characterized analytically and boils down to the intuitive static condition (7). This condition can be easily manipulated into mismatch indexes –measuring the distance between the actual and optimal allocation– that can be estimated using micro data. In the context of the recent U.S. experience, these indexes can help quantifying how much of the observed rise in unemployment is due to increased mismatch.

The transparency of our approach must be traded off against two drawbacks. First, some of the impediments to labor mobility, in particular moving and retraining costs, would be part of the physical

environment in a planner's problem and will likely lead to a lower mismatch relative to what we measure. Therefore, in this respect our approach should be thought of as a measurement device that (for a given level of disaggregation) delivers an *upper bound for the level of mismatch unemployment*.

Second, our methodology offers a measurement tool for mismatch unemployment, but does not fully get at the question of why unemployed workers are misallocated. Answering this question would require solving an equilibrium model incorporating all the potential sources of limited labor mobility across sectors. Within our approach, we can still learn about the deep sources of mismatch by examining how mismatch varies as we use different definitions of sector (occupation, industry, location, education).

3 Mismatch index and counterfactual unemployment

We now derive, from the optimality condition (7), an index measuring the severity of labor market mismatch. From this point onward we must state an additional assumption, well supported by the data as we show below: the individual-market matching function $m(u_i, v_i)$ is Cobb-Douglas, i.e.,

$$h_{it} = \Phi_t \phi_{it} v_{it}^{\alpha} u_{it}^{1-\alpha}, \tag{11}$$

where h_{it} are hires in sector i at date t, and $\alpha \in (0,1)$ is the vacancy share. To fix ideas, we begin with the case where there is no heterogeneity in ϕ, z or δ across markets, and then we move to the cases with heterogeneity. We then describe how to use these indexes to construct counterfactual experiments that show how much of the recent rise in U.S. unemployment is due to mismatch.

3.1 Mismatch index

The $\mathcal{M}_t^h$ index. We begin by assuming that the only sectoral-level heterogeneity is in the number of vacancies. Summing hires across markets by using (11), the aggregate numbers of hires can be expressed as:

$$h_t = \Phi_t v_t^{\alpha} u_t^{1-\alpha} \cdot \left[\sum_{i=1}^{I} \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} \right]. \tag{12}$$

The first term in the RHS of (12) denotes the highest number of new hires h_t^* that can be achieved under the optimal allocation of the u_t unemployed workers where market tightness (v_{it}/u_{it}) is equated (to its aggregate value v_t/u_t across sectors. Therefore, we can naturally define the mismatch index

$$\mathcal{M}_t^h = 1 - \frac{h_t}{h_t^*} = 1 - \sum_{i=1}^{I} \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha}. \tag{13}$$

The index $\mathcal{M}_t^h$ measures the fraction of hires lost in period t because of misallocation.[8] It answers the question: if the planner had u_t available unemployed workers and used its optimal allocation rule, how

[8] Expressed in terms of the observed hires, this fraction is $\mathcal{M}_t^h / \left(1 - \mathcal{M}_t^h\right)$.

many additional jobs would it create? These additional hires are created because, by better allocating the unemployed, the planner increases the aggregate job-finding rate. Using (13), one can rewrite the aggregate matching function as $h_t = \left(1 - \mathcal{M}_t^h\right) \Phi_t v_t^\alpha u_t^{1-\alpha}$ which makes it clear that higher mismatch lowers the efficiency of the aggregate technology and reduces the aggregate job-finding rate.

It is useful to discuss some properties of this index. First, it is easy to see that $\mathcal{M}_t^h \leq 1$. To show that $\mathcal{M}_t^h \geq 0$, note that

$$1 - \mathcal{M}_t^h = \frac{1}{v_t^\alpha u_t^{1-\alpha}} \sum_{i=1}^{I} \left(v_{it}\right)^\alpha \left(u_{it}\right)^{1-\alpha} \leq \frac{1}{v_t^\alpha u_t^{1-\alpha}} \left(\sum_{i=1}^{I} v_{it}\right)^\alpha \left(\sum_{i=1}^{I} u_{it}\right)^{1-\alpha} = 1,$$

where the $\leq$ sign follows from Hölder's inequality. Second, the $\mathcal{M}_t^h$ 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. Finally, the mismatch index is increasing in the level of disaggregation (i.e., the number of sectors). To see this, 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 $\mathcal{M}_I^h$ be the mismatch index over the I sectors and $\mathcal{M}_{IJ}^h$ be the one over the $I \times J$ sectors. Rearranging equation (13),

$$\begin{aligned}
1 - \mathcal{M}_I^h &= \frac{1}{v^\alpha u^{1-\alpha}} \sum_{i=1}^{I} \left(v_i\right)^\alpha \left(u_i\right)^{1-\alpha} \\
&= \frac{1}{v^\alpha u^{1-\alpha}} \sum_{i=1}^{I} \left(\sum\nolimits_{j=1}^{J} v_{ij}\right)^\alpha \left(\sum\nolimits_{j=1}^{J} u_{ij}\right)^{1-\alpha} \\
&= \frac{1}{v^\alpha u^{1-\alpha}} \sum_{i=1}^{I} \left(\sum\nolimits_{j=1}^{J} \left(v_{ij}^\alpha\right)^{\frac{1}{\alpha}}\right)^\alpha \left(\sum\nolimits_{j=1}^{J} \left(u_{ij}^{1-\alpha}\right)^{\frac{1}{1-\alpha}}\right)^{1-\alpha} \\
&> \frac{1}{v^\alpha u^{1-\alpha}} \sum_{i=1}^{I} \sum_{j=1}^{J} v_{ij}^\alpha u_{ij}^{1-\alpha} = 1 - \mathcal{M}_{IJ}^h,
\end{aligned}$$

where the last line uses Hölder's inequality. This last result suggests that every statement about the role of mismatch should be qualified with respect to the degree of sectoral disaggregation used. For this reason, we perform our empirical analysis at different levels of aggregation, whenever possible.

The $\mathcal{M}_{\phi t}^h$ index. Suppose now that labor markets differ in their frictional parameter ϕ_{it}. From equation (7), rearranging the optimality condition dictating how to allocate unemployed workers between market i and market j at date t, we obtain

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

Summing across j's yields

$$u_{it}^{*}=\phi_{it}^{\frac{1}{\alpha}}\cdot\left(\frac{v_{it}}{\sum_{i=1}^{I}\phi_{it}^{\frac{1}{\alpha}}v_{it}}\right)\cdot u_{t}. \tag{14}$$

The optimal aggregate number of hires is

$$h_{t}^{*}=\Phi_{t}v_{t}^{\alpha}u_{t}^{1-\alpha}\left[\sum_{i=1}^{I}\phi_{it}\left(\frac{v_{it}}{v_{t}}\right)^{\alpha}\left(\frac{u_{it}^{*}}{u_{t}}\right)^{1-\alpha}\right]. \tag{15}$$

Substituting the planner's allocation rule (14) in equation (15) , the total number of optimal new hires is $h_{t}^{*}=\Phi_{t}\bar{\phi}_{t}v_{t}^{\alpha}u_{t}^{1-\alpha}$, where

$$\bar{\phi}_{t}=\left[\sum_{i=1}^{I}\phi_{it}^{\frac{1}{\alpha}}\left(\frac{v_{it}}{v_{t}}\right)\right]^{\alpha} \tag{16}$$

is a CES aggregator of the market-level matching efficiencies weighted by their vacancy share. Similarly, we can define the total number of observed new hires as

$$h_{t}=\Phi_{t}v_{t}^{\alpha}u_{t}^{1-\alpha}\left[\sum_{i=1}^{I}\phi_{it}\left(\frac{v_{it}}{v_{t}}\right)^{\alpha}\left(\frac{u_{it}}{u_{t}}\right)^{1-\alpha}\right], \tag{17}$$

and, hence, the counterpart of (13) in the heterogeneous matching efficiency case becomes

$$\mathcal{M}_{\phi t}^{h}=1-\frac{h_{t}}{h_{t}^{*}}=1-\sum_{i=1}^{I}\left(\frac{\phi_{it}}{\bar{\phi}_{t}}\right)\left(\frac{v_{it}}{v_{t}}\right)^{\alpha}\left(\frac{u_{it}}{u_{t}}\right)^{1-\alpha}. \tag{18}$$

The index in (18) is similar to the index (13) derived for the homogeneous markets case, except for the adjustment term in brackets which equals one when there is no heterogeneity in matching efficiencies, i.e., $\phi_{it}=1$ for all i. This term corrects the index for the fact that the planner may want to allocate a share of unemployed workers larger than the vacancy share in market i when its matching efficiency ϕ_{it} is higher than the weighted average $\bar{\phi}_{t}$.

The $\mathcal{M}_{xt}^{h}$ index. We now describe how to compute mismatch indexes when labor markets differ in their level of productivity. Our derivations below apply to the model of Section 2.3.1. It is immediate that to obtain the mismatch index for the economy of Section 2.3.2 it suffices substituting z_{it} with the term $\frac{Z_{t}^{\eta_{i}}}{1-\beta(1-\delta_{t})\left(\exp\left(\eta_{i}(\eta_{i}-1)\frac{\sigma_{\varepsilon}}{2}\right)\right)}$, and to obtain the mismatch index in the economy with heterogeneous destruction rates of Section 2.3.3, it suffices substituting z_{it} with $\frac{z_{it}}{(1-\beta(1-\delta_{t})(1-\delta_{it}))}$.

It is useful to define "overall market efficiency" as the product $x_{it}\equiv z_{it}\phi_{it}$ of productive and matching efficiency of sector i. The optimality condition dictating how to allocate unemployed workers between market i and market j is:

$$\frac{v_{it}}{u_{it}^{*}}=\left(\frac{x_{jt}}{x_{it}}\right)^{\frac{1}{\alpha}}\cdot\frac{v_{jt}}{u_{jt}^{*}}. \tag{19}$$

The optimal number of hires that can be obtained by the planner allocating the available unemployed workers across sectors is still given by equation (15) . Substituting the optimality condition (19) in equation (15) , the optimal number of new hires is $h_t^* = \Phi_t \bar{\phi}_{xt} v_t^{\alpha} u_t^{1-\alpha}$, where

$$\bar{\phi}_{xt} = \bar{x}_t \cdot \frac{\sum_{i=1}^{I} \left(\frac{1}{z_{it}}\right) x_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t}\right)}{\sum_{i=1}^{I} x_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t}\right)},$$

and

$$\bar{x}_t = \left[\sum_{i=1}^{I} x_{it}^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t}\right)\right]^{\alpha} \tag{20}$$

is a CES aggregator of the market-level overall efficiencies weighted by their vacancy share. Comparing equations 3.1 and 16 reveals that if z_{it} is constant across markets, $\bar{\phi}_{xt} = \bar{\phi}_t$. Since total new hires are given by (17), we obtain the mismatch index

$$\mathcal{M}_{xt}^{h} = 1 - \sum_{i=1}^{I} \left(\frac{\phi_i}{\bar{\phi}_{xt}}\right) \left(\frac{v_{it}}{v_t}\right)^{\alpha} \left(\frac{u_{it}}{u_t}\right)^{1-\alpha}, \tag{21}$$

which measures the fraction of hires lost because of mismatch at date t.

In what follows, we will also use the notation $\mathcal{M}_{zt}^{h}$ to denote mismatch indexes for an economy where the only source of heterogeneity, beyond vacancies, is productivity and $\mathcal{M}_{\delta t}^{h}$ for an economy where the only source of heterogeneity is job destruction rates.

Measurement. Suppose one can access longitudinal data on unemployment $\{u_{it}\}$ and vacancies $\{v_{it}\}$ for various sectors at different dates. Then, using data on hires $\{h_{it}\}$, from equation (11) one can consistently estimate the (common, across sectors) vacancy share α as well as the vector of sector-specific matching efficiencies $\{\phi_i\}$ under the assumption that the latter are constant over the sample period used for the estimation. Data on labor productivity and job destruction rates by sector can be used to measure $\{z_{it}, \delta_{it}\}$. These are all the necessary ingredients to construct time series for all the misallocation indexes defined above. Section 4 below illustrates this measurement step in detail.

3.2 Counterfactual unemployment

The key question of interest is: how much smaller would the recent rise in the U.S. unemployment rate have been without any mismatch between unemployment and vacancies across sectors? To answer this question, we construct a counterfactual unemployment rate based on the planner's optimal cross-sectional unemployment allocation rule.

To simplify the notation, consider the baseline version of the model and the corresponding index $\mathcal{M}_t^h$. As explained in Section 3.1, the actual aggregate job-finding rate in the economy at date t is

$$f_t = \frac{h_t}{u_t} = \left(1 - \mathcal{M}_t^h\right) \cdot \Phi_t \cdot \left(\frac{v_t}{u_t}\right)^{\alpha}.$$

Let u_t^* be counterfactual unemployment under the planner's allocation rule. From the discussion of equation (12) recall that the optimal number of hires in period t when u_t^* unemployed workers are available to be allocated across sectors is $\Phi_t v_t^{\alpha} (u_t^*)^{1-\alpha}$. Therefore, the optimal job-finding rate (without mismatch) is

$$f_t^* = \frac{h_t^*}{u_t^*} = \Phi_t \cdot \left(\frac{v_t}{u_t^*}\right)^{\alpha} = f_t \cdot \frac{1}{\left(1 - \mathcal{M}_t^h\right)} \left(\frac{u_t}{u_t^*}\right)^{\alpha}, \tag{22}$$

where the last term in this equation shows that, in the counterfactual, the aggregate job-finding rate is higher than the observed one (i) because of the absence of mismatch and (ii) because lower unemployment $(u_t^* < u_t)$ increases the probability of meeting a vacancy for the job-seekers. Given an initial value for u_0^*, the counterfactual frictional unemployment rate can be obtained by iterating forward the equation

$$u_{t+1}^* = s_t + \left(1 - s_t - f_t^*\right) u_t^*, \tag{23}$$

where s_t is the separation rate.[9] By comparing the dynamics of the data (u_t) to those of the counterfactual unemployment (u_t^*), one can gauge the role of mismatch in the labor market.

This 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 (23). This approach is consistent with the theoretical model where vacancy creation and separations are exogenous to the planner. In the models of Section 2.3 where labor force participation and separation choices are endogenous, the planner could, potentially, take (separation and labor force) decisions different from the corresponding equilibrium outcomes –those we observe in the data. Therefore, in these cases u_t^* cannot be interpreted as "planner's unemployment rate" but it should be strictly interpreted as the counterfactual unemployment rate under the planner's allocation rule of idle workers across sectors —abstracting from other possible discrepancies between the planner's separation and labor force participation choices and the corresponding equilibrium outcome.

[9] We calculate the aggregate separation rate s_t and the job-finding rate f_t using the methodology described in Shimer (2005). Consequently f includes transitions into nonparticipation as well as employment. We apply our correction to this total outflow rate and do not make a distinction between flows depending on their destination. As Shimer (2007b) shows in his Figure 4, the ratio of unemployment-to-employment flow rate to the unemployment-to-nonparticipation flow rate is very stable over the business cycle. Thus, our counterfactual gives us an upper bound on the effect of mismatch on the job-finding rate but does not cause a cyclical bias on the effect of mismatch on the unemployment rate.

4 Data and Sectoral Matching Functions

We begin this section by describing the data sources. Next we analyze the issue of specification and estimation of the matching function.

4.1 Data Description

In our analysis, we focus on four major definitions of labor markets: the first is a broad industry classification; the second is an occupation classification, based on both 2-digit and 3-digit SOC's; the third is a geographic classification, based on U.S. states; finally, we also study mismatch within four skill categories, based on educational attainment.

As discussed in Section 3, our analysis requires detailed information about vacancies, hires, unemployment, productivity, and job destruction rates across different labor markets.

Vacancy and hire data at the industry level come from the Job Openings and Labor Turnover Survey (JOLTS) which provides survey-based measures of job openings and hires at a monthly frequency for seventeen industry classifications.[10] At the occupation, education and state level we use vacancy data from the Help Wanted OnLine (HWOL) dataset provided by The Conference Board (TCB). We describe these data in more detail below. With regard to the unemployed, we calculate unemployment counts from the CPS for the same industry, occupation, geography and education classifications that we use for vacancies.[11]

Computation of mismatch indexes with heterogenous productive and matching efficiency requires estimates of labor-market specific productivities, matching efficiencies, and shares of the matching function. We use various proxies for productivity, depending on data availability. At the industry level, we use data on gross output and employment from the EU KLEMS Growth and Productivity Accounts and compute sectoral productivity measures.[12] At the occupation level, we use average hourly earnings from the Occupational Employment Statistics (OES).[13] We recognize the fact that wage levels might be affected by reasons other than productivity like unionization rates, compensating differentials, etc. To 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.[14] Since the BED is quarterly, we assume that

[10] For more details on the JOLTS, see http://www.bls.gov/jlt/.

[11] 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.

[12] See http://www.euklems.net/ for details.

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

[14] http://www.bls.gov/bdm/

the destruction rate is the same for the three months corresponding to a specific quarter.[15] We have not attempted to compute job destruction rates for other definitions of sector.

The calculation of market-specific match efficiency parameters, ϕ_i, and vacancy share α is more involved. We describe its details below.

For unemployed workers, the CPS reports the industry and occupation of the worker's previous job, while for employed workers, the survey reports the industry and occupation of the current job. Ideally, we would like to count how many unemployed workers are *searching* for jobs in a particular sector and this number does not necessarily coincide with the number of workers whose last employment was in that sector. We attempt a correction when we compute mismatch by industry by exploiting the semi-panel dimension of the CPS. Since respondents in the CPS are interviewed for several consecutive months, given any two adjacent months, we can track unemployed workers who find new employment from one month to the next. Thus we can obtain two key facts about unemployed workers who find jobs: 1. the industry of the previous job prior to the workers unemployment spell; 2. the industry of the new job. We create annual transition rate matrices by aggregating monthly data and calculating a five year centered moving average for 2001-2010. We exclude individuals (unemployed and employed workers) who do not have an industry classification. In implementing this procedure, we follow Hobijn (2011). We then infer the number of job seekers in each industry using the method outlined in Appendix A2.

4.1.1 The online vacancy data

The Help Wanted OnLine (HWOL) dataset provided by The Conference Board (TCB) is a novel data series that covers the universe of online advertised vacancies posted on internet job boards or on newspaper online editions.[16] The HWOL data base started in May 2005 as a replacement for the Help-Wanted Advertising Index of print advertising maintained by TCB. It covers roughly 1,200 online job boards and provides detailed information about the characteristics of advertised vacancies for several million active ads each month. When the same ad for a given position is posted on multiple job boards, an unduplication algorithm is used that identifies unique advertised vacancies on the basis of the combination of company name, job title/description, city or State.

Each observation in the HWOL data base 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 full-time or part-time. The education level of the position, and the hourly and annual mean wage are imputed.[17] For a subset of ads we also observe the industry

[15] BED data get released on average with a three quarter delay. The latest BED release contains data for the fourth quarter of 2010. Consequently, our analysis with heterogenous destruction rates contains data up to December 2010.

[16] The data are collected for The Conference Board by Wanted Technologies.

[17] The education level is imputed by TCB based on BLS information on the education content of detailed 6-digit level occupations. We classify vacancies by education level using an algorithm that we describe in detail in Section 5.4 below.

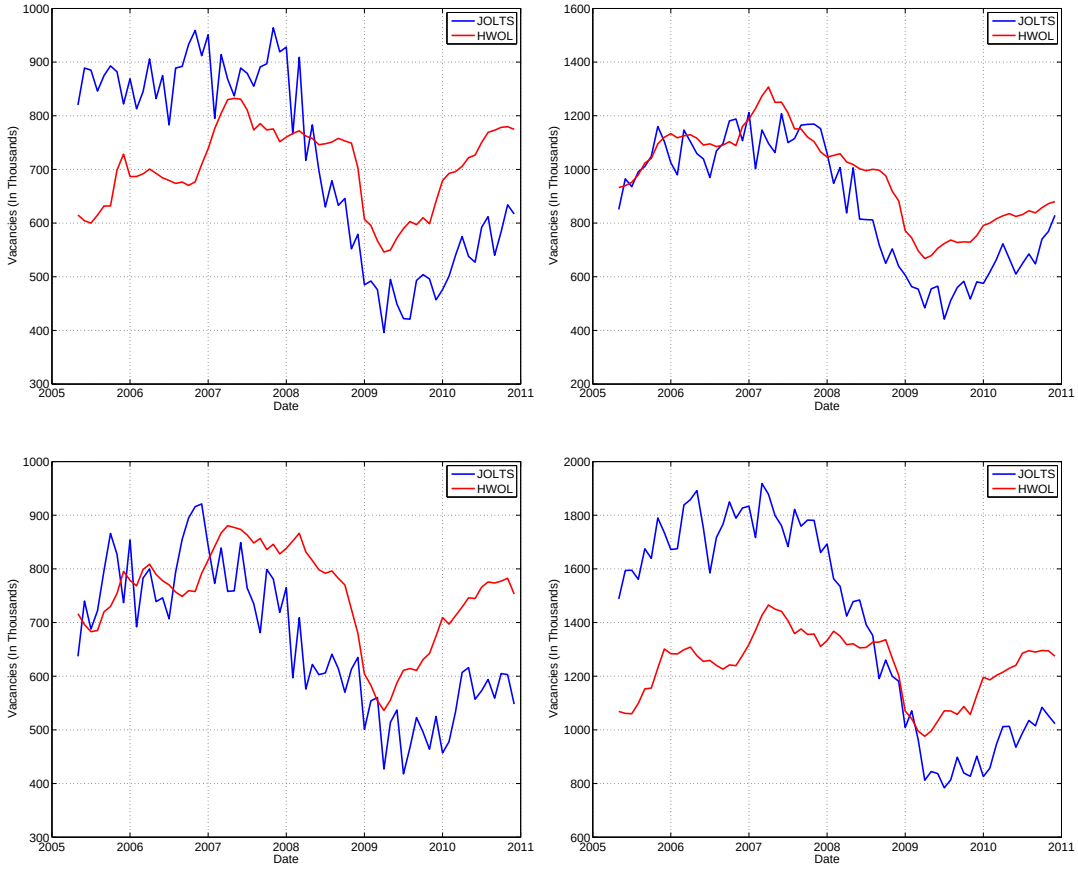

Figure 1: Comparison Between JOLTS and HWOL. Top-left panel: Midwest, Top-right panel: West, Bottom-left panel: Northeast, Bottom-right panel: South.

NAICS classification, the sales volume and number of employees of the company, and the advertised salary.

The aggregate trends from the HWOL data base are roughly consistent with those from the JOLTS data: in Figure 1 we plot JOLTS vacancies and HWOL ads by Census region. At the national level, the total count of active vacancies in HWOL is slightly below that in JOLTS until the beginning of 2008, and is above from 2008 onwards. This difference is most pronounced in the South, and may reflect the growing penetration of online job listings over time. The average difference between the two aggregate series is about 11% of the total. The correlation between the two aggregate series is very high, 0.91, indicating that the patterns over time are very similar.

The vast majority of online advertised vacancies is posted on a small number of job boards: about 70% of all ads appears on nine job boards;[18] about 60% is posted on only five job boards. It is worth mentioning some measurement issues in the HWOL data: first, as mentioned earlier, there seems to

Wages are imputed from BLS data on Occupational Employment Statistics (OES), based on the occupation classification.

[18]These are: "Absolutely Health Care", "Craigslist", "JOBcentral", "CareerBuilder", "Monster", "Yahoo!HotJobs", "Recruiter Networks", "Dice","DataFrenzy".

		CES	Cobb Douglas
σ	Point estimate	-0.074	
	95% Conf. Interval	$(-0.267, 0.081)$	-
α	Point estimate	0.512	0.532
	95% Conf. Interval	$(0.466, 0.551)$	$(0.514, 0.553)$
Φ	Point estimate	0.939	0.943
	95% Conf. Interval	$(0.922, 0.958)$	$(0.925, 0.962)$

Table 1: CES vs. Cobb Douglas

be a slight time trend in the time series for HWOL vacancies relative to JOLTS, perhaps reflecting the growing use of online job boards over time. This should not overly affect our indices given the very high correlation between the two series.[19]

Secondly, the dataset records one vacancy per ad. There is a small number of cases in which multiple positions are listed, but the convention of one vacancy per ad is used for simplicity. Finally, there are some cases in which multiple locations (counties within a state) are listed in a given ad for a given position. However, this is not an issue for our analysis since we focus on states as the smallest unit of geographic analysis at present.

Currently, we use HWOL data to construct mismatch indexes by 2-digit and 3-digit occupation, by state, as well as within education groups. Given the richness of detail of the vacancy information contained in HWOL, the limitations in constructing finer mismatch indexes arise from the unemployment side because of the relatively small size of the CPS. In related work, we are using job seeker data from public career centers in individual states to conduct a more detailed analysis of mismatch for selected states.[20]

4.2 Matching function estimation

We start by showing that a matching function with unit elasticity is a reasonable representation of the hiring process at the sectoral level. Using the JOLTS data for the 2-digit definition of industries and the period December 2000-December 2010, we estimate the parameters of the following CES

[19] In future work we plan to perform some robustness checks restricting the sample to a subset of job boards that have been more stable over time, to mimic the JOLTS series more closely.

[20] To give an idea of the level of detail represented by different levels of aggregation, there is a single two-digit category corresponding to "Healthcare Practitioners and Technical Occupations". This category is sub-divided into "Health Diagnosing and Treating Practitioners", "Health Technologists and Technicians" and "Other" at the three-digit level. At the six-digit level, one sees such detail as "Occupational Therapists", "Physical Therapists", or "Speech-Language Pathologists". More detail can be found at http://www.bls.gov/soc/2000/soc_j0a0.htm

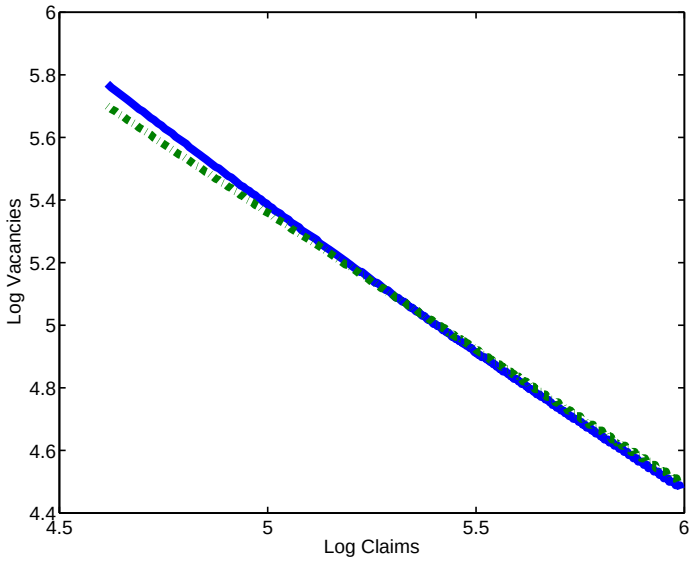

Figure 2: Iso-matching curves for CES (Solid) and Cobb-Douglas (Dashed)

matching function via minimum distance:[21]

$$\ln\left(\frac{h_{it}}{u_{it}}\right) = \ln\Phi + \frac{1}{\sigma}\ln\left[\alpha\left(\frac{v_{it}}{u_{it}}\right)^{\sigma} + (1-\alpha)\right]. \tag{24}$$

Recall that $\sigma \in (-\infty, 1)$ with $\sigma = 0$ in the Cobb-Douglas case.[22] As the left column of Table 1 indicates, we find that $\hat{\sigma} = -0.074$ implying an elasticity around 0.93, hence only slightly smaller than the Cobb-Douglas benchmark. Moreover, $\hat{\sigma}$ is not significantly different than zero at the 5% significance level. The right panel of Table 1 reports estimation results for the Cobb-Douglas case (i.e., imposing the constraint $\hat{\sigma} = 0$). The results indicate that there is no statistically significant difference in the estimates $(\hat{\alpha}, \hat{\Phi})$ between the CES and the Cobb-Douglas case; therefore the latter specification is a good approximation for the matching function at this level of aggregation. Figure 2 plots the iso-matching curves for the CES and the Cobb-Douglas specifications over the empirical range of vacancies and unemployment, demonstrating the closeness of the two specifications. In light of this finding, and given the analytical convenience of the unit elasticity benchmark, we restrict σ to be zero and use a Cobb-Douglas matching function throughout the paper.

The next step is to estimate the parameters of the matching function that are required for comput-

[21]Note that JOLTS reports vacancies and hires on the last day of the month and the CPS reports the number of unemployed during the survey week, which is the week containing the 12th day of the month. To be consistent with the timing of the measurement of flows and stocks, we use unemployment and vacancy stocks in month $t-1$ and hires in month t in all regressions.

[22]We use simulated annealing to minimize the minimum distance criterion to ensure that we obtain a global minimum. 95% confidence intervals are computed via bootstrap methods.

	Truncated Sample		Full Sample	
	OLS	Fixed Effects	OLS	Fixed Effects
Aggregate	0.611	-	0.797	-
	(0.018)	-	(0.014)	-
Aggregate (Quadratic Time Trend)	0.691	-	0.673	-
	(0.026)	-	(0.011)	-
Industry	0.402	0.504	0.529	0.671
	(0.007)	(0.010)	(0.009)	(0.012)
Industry (Quadratic Time Trend)	0.385	0.500	0.445	0.556
	(0.013)	(0.012)	(0.009)	(0.008)

Note: Standard errors in parenthesis. Truncated sample: Dec 2000 to Dec 2007. Full sample: Dec 2000-Dec 2010.

Table 2: Estimates of the vacancy share α

ing mismatch indexes. We start by estimating an aggregate matching function of the form

$$\ln\left(\frac{h_t}{u_t}\right) = \ln \Phi_t + \alpha \ln\left(\frac{v_t}{u_t}\right)$$

where h_t is the number of matches, u_t is unemployment and v_t in the number of vacancies in month t. We use hires from the JOLTS as our measure of matches.[23] Vacancies come from the JOLTS and aggregate unemployment numbers come from the CPS. The first row of Table 2 reports estimates of α for two sample periods. The estimate for α is 0.797 if we use our full sample which spans December 2000 to December 2010. When we constrain the sample to pre-recession data (December 2000 to December 2007), the estimate for α is lower at 0.611. As we have discussed earlier, there is potentially some time variation in Φ. To capture the time variation in Φ, we run a similar regression with a quadratic time trend: the results are reported in the second row of Table 2. With the quadratic time trend, estimates of α are much closer for the full sample and the pre-recession sample at around 0.67-0.69.

In addition to the aggregate regressions, we also exploit industry-level data on hiring, vacancies and unemployment and estimate the following regression

$$\ln\left(\frac{h_{it}}{u_{it}}\right) = \ln \Phi_t + \ln \phi_i + \alpha \ln\left(\frac{v_{it}}{u_{it}}\right) \tag{25}$$

for both our full and pre-recession samples. We constrain Φ_t to be the same across sectors and allow for a quadratic time trend to control for time variation. The results of the estimation without fixed effects are reported in the last two rows of Table 2, in the columns labeled "OLS". The estimates of α are lower than the ones estimated by the aggregate regression varying between 0.38 and 0.53. As in

[23] An alternative is to use the unemployment outflow rate or the unemployment to employment transition rate. Here we focus on the direct measure of industry-specific hires provided by the JOLTS.

Sector	ϕ
Mining	1.49
Arts	1.45
Construction	1.40
Accommodations	1.33
Retail	1.27
Professional Business Services	1.26
Real Estate	1.24
Wholesale	1.07
Other	1.00
Transportation and Utilities	0.98
Health	0.83
Manufacturing - Nondurables	0.82
Education	0.81
Government	0.77
Finance	0.76
Manufacturing - Durables	0.72
Information	0.66

Table 3: Industry-specific matching efficiencies

the case of aggregate regressions, allowing for time variation lowers the estimate of α. The estimation results with industry fixed-effects in matching efficiency are reported in the last two rows of Table 2, in the columns labeled "Fixed Effects". In these cases, estimates of α vary between 0.50-0.67 with higher estimates when we use the full sample.

To sum up, depending on the specification, estimates of α vary in the range 0.38-0.80. To pick a value for α, we are guided by two criteria. First, to be consistent with the theory, in the aggregate regressions one should favor models with time-varying aggregate matching efficiency, and in the sectoral regressions models with fixed effects. Such models are also more flexible. Second, it is easy to see how a structural break in Φ which lowers the intercept of (25) at a time where market tightness is low could severely bias upward the elasticity α. As a result, we favor estimates of α based on the truncated pre-recession sample. Based on these two criteria, the range narrows to 0.50-0.69, which invites a choice of $\alpha = 0.60$.[24]

Estimation of (25) also provides us with sector-specific estimates of match efficiency (ϕ_i). These estimates are reported in Table 3, using the pre-recession sample.[25] Industry-specific match efficiency estimates (ϕ_i) vary considerably and are between 0.63 to 1.5. Education, health, finance, and information stand out as low-efficiency sectors while construction stands out as a high efficiency sector. One interpretation of these differences is that general skill labor markets have the highest (ϕ_i) and

[24]Estimates of α using HWOL vacancy data are roughly consistent with the ones obtained using JOLTS.

[25]Our mismatch analysis with heterogeneous match efficiency uses these estimates, based on the pre-recession sample. Using the full sample does not change the empirical findings much quantitatively.

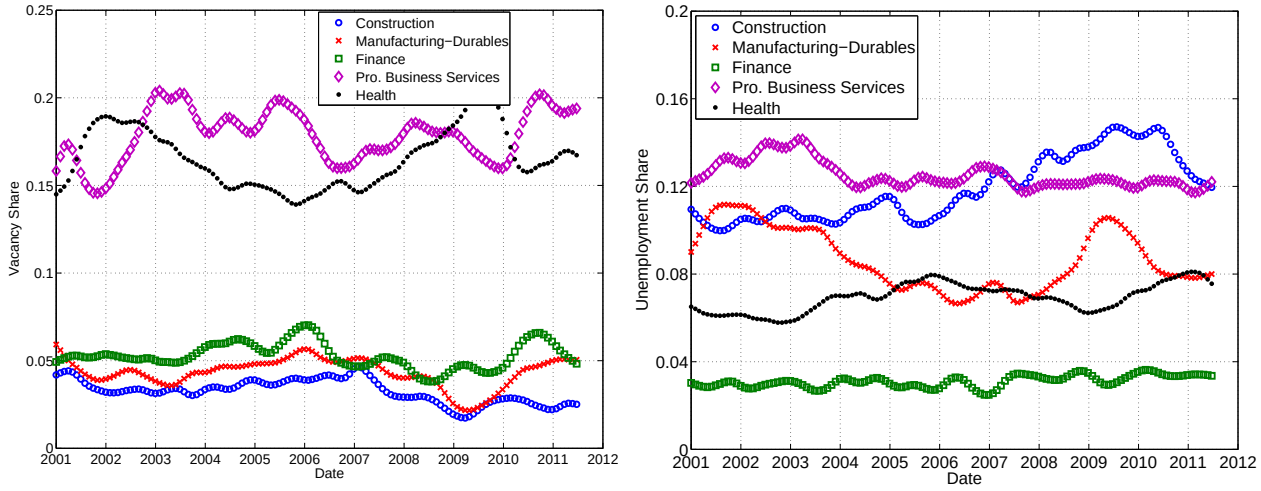

Figure 3: Vacancy and unemployment share by selected industry.

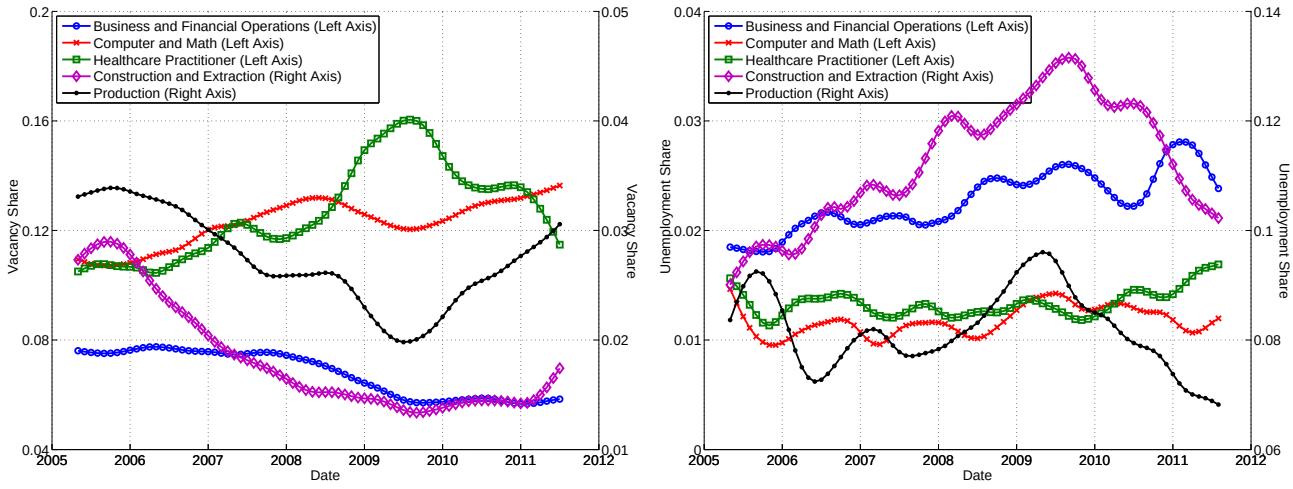

Figure 4: Vacancy and unemployment shares by selected occupations.

specialized skill labor markets the lowest (ϕ_i). High efficiency might also be an outcome of different hiring practices in different industries (e.g., informal referrals). Finally, heterogeneity in measured match efficiency could derive from different degrees of underreporting of vacancies across sectors, as discussed in Davis, Faberman, and Haltiwanger (2010). In Appendix A3, we show that ϕ_i is proportional to the fraction of underreported vacancies and that the mismatch index $\mathcal{M}^h_{\phi t}$ is robust to such measurement error.

4.3 A First Look At Mismatch

As a preliminary investigation, it is useful to examine the vacancy and unemployment shares of different sectors, occupations and geographic areas since these statistics are inputs into our mismatch indexes. If vacancy and unemployment shares of different labor markets do not vary over time, there

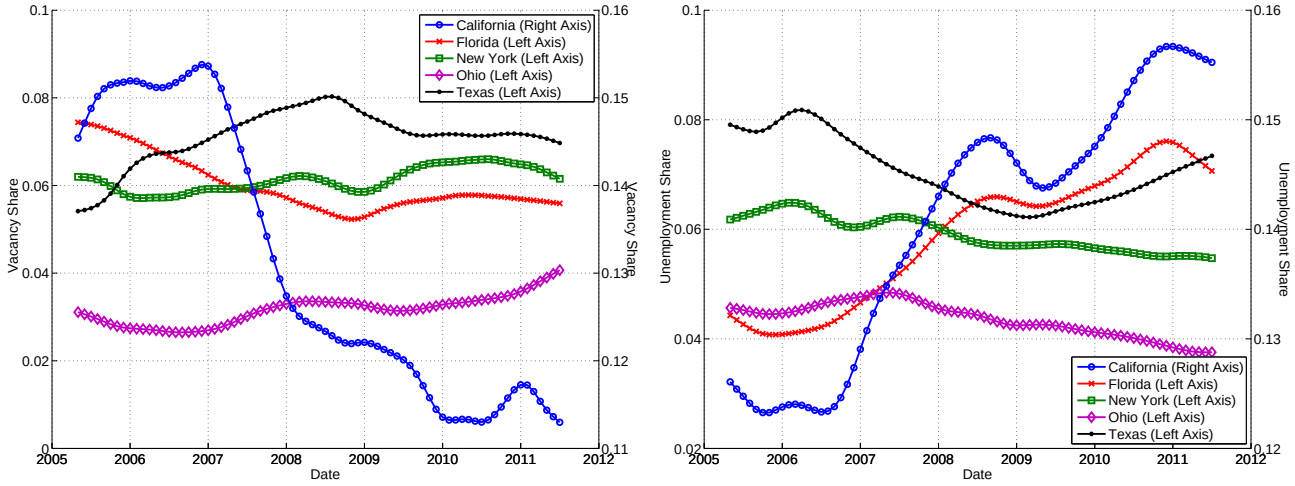

Figure 5: Vacancy and unemployment shares by selected states.

is little room for mismatch to play an important role in the increase in the unemployment rate. To examine this issue, we first plot the vacancy and unemployment shares for a selected set of industries using the JOLTS definition. As Figure 3 shows, 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. Interestingly starting from 2010, unemployment and vacancy shares of sectors began to normalize and almost went back to their pre-recession levels with the exception of the construction sector. The vacancy share of the construction sector remains well below its pre-recession level.

Now turning to the HWOL data, we plot the vacancy and unemployment shares for a selected set of occupations and U.S. states. Figure 4 shows the unemployment and vacancy shares of selected 2-digit occupations. 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 which experienced a decline in their vacancy shares and an increase in their unemployment shares. Concurrently, vacancy shares of health-care practitioner and computer and math occupations went up. Starting from 2010, similar to the JOLTS data, unemployment and vacancy shares began to normalize. For instance, for production occupations the vacancy share almost went back to their pre-recession levels, while for construction and extraction occupations the vacancy share is still considerably different from its pre-recession levels. These patterns suggest that skill mismatch measured at the occupation level may have increased during the recession, but started to revert back as the recovery in the labor market began.

Figure 5 shows the behavior of vacancy and unemployment shares for a selection of U.S. states.

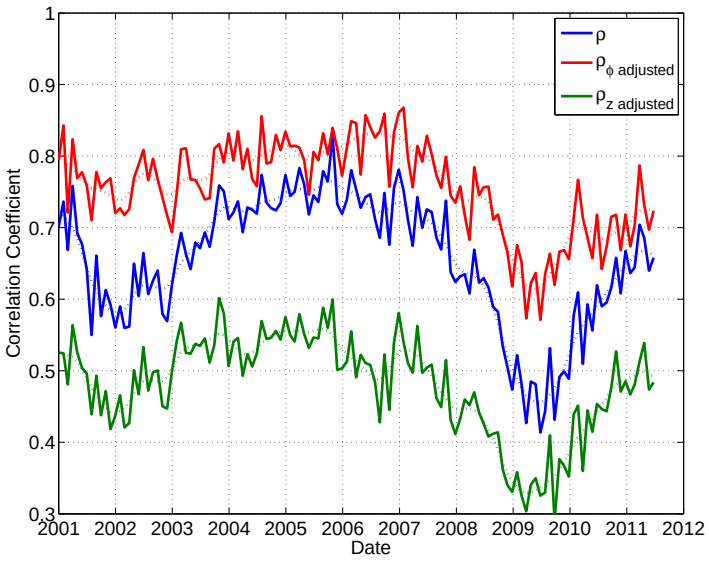

Figure 6: Correlation coefficient between u and v shares.

California and Florida were hit hard by the recession, as reflected by the decline in their vacancy shares and the notable increase in their unemployment shares. As one might expect, California experienced a drastic deterioration of labor market conditions: California's vacancy share went down from over 15% to 11% and its unemployment share went up by 4 percentage points, from around 12% to almost 16%. New York, Ohio and Texas fared relatively better. Unemployment and vacancy shares still seem quite different from their pre-recession levels: this may be potentially due to a differential geographic impact of the recession as well as to other long-run differences in regional trends.

5 Empirical results

This section collects the results of our empirical analysis of mismatch by industry, occupation, U.S. state, and education. We also perform the counterfactual exercises described in Section 3.2.

5.1 Industry-level mismatch

We present a first set of results on mismatch unemployment across the 17 industries classified in JOLTS. From our definition of mismatch in the labor market, 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 6 plots the time series of this correlation coefficient across industries over the sample period. In particular, we report three different correlation coefficients motivated by the definitions of

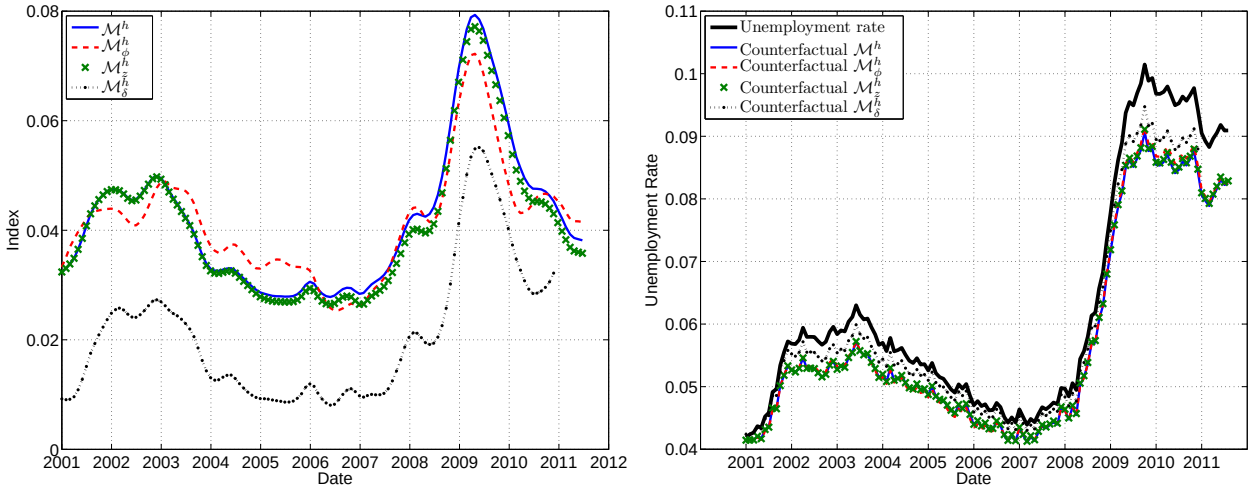

Figure 7: Mismatch index $\mathcal{M}_t^h$ by industry (left panel) and corresponding counterfactuals (right panel).

the mismatch indexes we derived in Section 3: 1. ρ: between (u_{it}/u_t) and (v_{it}/v_t); 2. ρ_ϕ: between (u_{it}/u_t) and $(\phi_i/\bar{\phi}_t)^{\frac{1}{\alpha}}(v_{it}/v_t)$, and 3. ρ_z: between (u_{it}/u_t) and $(z_i/\bar{z}_t)^{\frac{1}{\alpha}}(v_{it}/v_t)$. All three series behave very similarly. The basic correlation coefficient (ρ) drops from 0.75 in early 2006 to 0.45 in mid 2009 and recovers thereafter, indicating a rise in mismatch during the recession.

Left panel of figure 7 plots the $\mathcal{M}_t^h$ indexes in their various versions described in Section 3: the baseline index, $\mathcal{M}_t^h$; the one adjusted for heterogeneity in matching efficiency, $\mathcal{M}_{\phi t}^h$; the one adjusted for heterogeneity in productivity $\mathcal{M}_{zt}^h$; and the one modified to account for heterogeneous job destruction rates, $\mathcal{M}_{\delta t}^h$. 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 1 to 3 percent per month, depending on the index used. At the end of the recession, in mid 2009, it had increased to roughly 5-8 percent per month, and it has since dropped again. To sum up, the different variations of $\mathcal{M}_t^h$ all indicate a 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? Right panel of figure 7 contains the observed unemployment rate series and the counterfactual unemployment rates constructed following the strategy of Section 3.2 for each of our indexes. Figure 8 shows mismatch unemployment (i.e., the difference between the actual and the counterfactual unemployment rates) at the industry level for the 2006-2011 period, as implied by the various indexes. The main finding is that worsening mismatch across industries explains between 0.6 and 0.8 percentage points of the rise in U.S. unemployment from 2006 to its peak, depending on the index used, i.e., at most 16 percent of the increase.[26] Thus, at its peak, mismatch unemployment across industries con-

[26]Note that the average unemployment rate was 4.6% in 2006 and 10.1% at its peak in October 2009, indicating a 5.5 percentage point increase. Throughout the paper we compare the average of 2006 with the unemployment peak (October

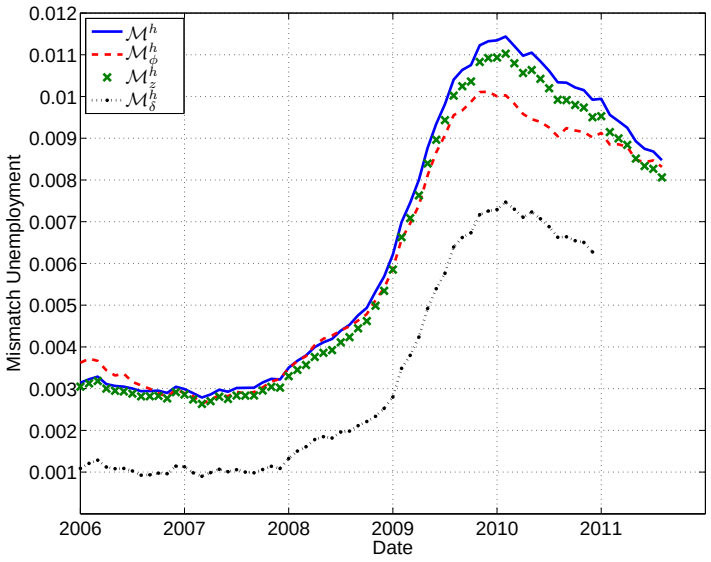

Figure 8: Mismatch unemployment measured at the industry level, 2006-2011.

tributed at most about 0.8 percentage points to the increase in the unemployment rate. It has declined starting in 2010, but still remains above its pre-recession levels.

5.1.1 Industry-level Mismatch with Adjusted Unemployment Counts

The empirical results we presented above assume that each unemployed worker is searching in the same industry as the one where she was last employed. We relax this assumption and infer the number of job seekers in each industry using the method outlined in Appendix A1.3 and the data described in Section 4. The left panel of Figure 9 shows the mismatch index $\mathcal{M}_t^h$ calculated using the adjusted unemployment counts as well the baseline $\mathcal{M}_t^h$ index. The adjustment causes the level of the index to increase by about 0.01 to 0.03 during the sample period. We also compute the counterfactual unemployment rate corresponding to the adjusted index as shown in the right panel of Figure 9. Not surprisingly, the counterfactual unemployment rate implied by the adjusted counts is lower than our baseline case, however in terms of accounting for the increase in the unemployment rate both indexes have remarkably similar quantitative implications. According to both indexes, 0.8 percentage points of the roughly 5 percentage point rise in U.S. unemployment is due to industry-level mismatch.

5.2 Occupational-level mismatch

We now present our results on mismatch unemployment across two- and three-digit occupations based on HWOL job advertisement and CPS unemployment data. Note that the HWOL ads data begin in May 2005 and the latest observation is July 2011.

2009) when we discuss the role of mismatch in the increase in the unemployment rate.

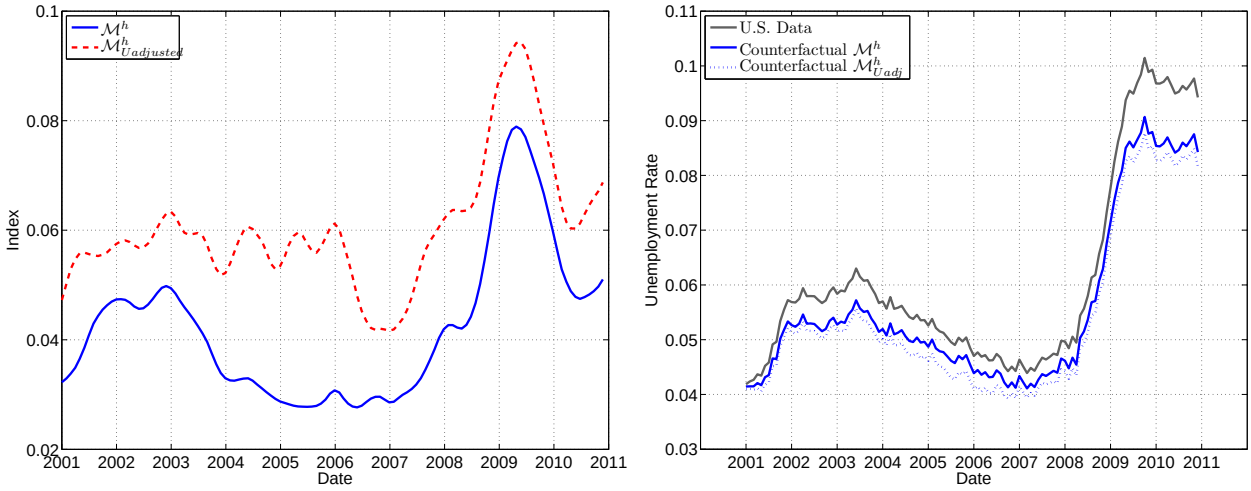

Figure 9: Mismatch index $\mathcal{M}_t^h$ by industry with unadjusted and adjusted unemployment counts (left panel) and corresponding counterfactuals (right panel).

5.2.1 2-digit occupations

Figure 10 plots the correlation between vacancy and unemployment shares across 2-digit SOC's. As for the industry-level analysis, we document a significant decline in the correlation, from 0.5 to about 0.25 between 2006 and 2009. This fall in the correlation is the counterpart of an increase in mismatch indexes. Figure 11 plots $\mathcal{M}_t^h$ and $\mathcal{M}_{zt}^h$ indexes (left panel) and the resulting counterfactual unemployment analysis (right panel) for 2-digit SOC's. Both $\mathcal{M}_t^h$ and $\mathcal{M}_{zt}^h$ rise by almost 0.04 over the same period, i.e., the fraction of monthly hires lost because of occupational mismatch grew from 0.1 to 0.14 over that period, a larger rise compared to the industry-level index. Moreover, the level of the index is substantially higher. We then calculate the mismatch unemployment rate (i.e., the difference between the actual and the counterfactual unemployment rates). As seen in Figure 12 around 1.4 percentage points (or around one quarter) of the recent surge in US unemployment can be attributed to occupational mismatch measured at the 2-digit occupation level.

5.2.2 3-digit occupations

Figure 11 plots the baseline $\mathcal{M}_t^h$ index (left panel) and the resulting counterfactual unemployment analysis (right panel) for 3-digit SOC's.[27] As the left panel of Figure 11 shows, the $\mathcal{M}_t^h$ index is significantly higher than for 2-digit occupations, but shows a very similar pattern over time. The increase in the index between 2006 and 2009 is about 0.04, indicating a similar increase in the fraction of hires lost due to mismatch as for 2-digit occupations. Looking at the counterfactual exercise, the fraction

[27]There are 93 three-digit SOC's. We exclude 3-digit SOC's that never exhibit more than 10 observations in the CPS unemployment counts. These small cells account for 60% of the 3-digit SOC's, but represent only 15.6% of unemployed workers in the CPS.

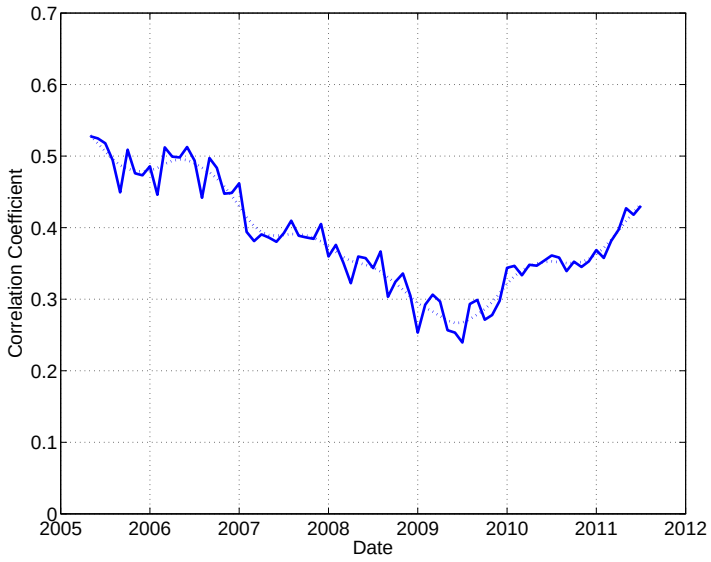

Figure 10: Correlation coefficient between vacancy and unemployment shares across occupations.

of unemployment due to occupational mismatch is significantly higher throughout the sample period at this finer level of aggregation than at the 2-digit level. However, as Figure 12 shows, the portion of the increase in unemployment attributable to mismatch is around 1.7 percentage points which is roughly comparable to our finding at the two-digit level. Thus, performing the analysis at different levels of aggregation yields similar results in terms of the *change* in the fraction of unemployment attributable to occupational mismatch.

5.3 Geographical mismatch

We now turn to geographical mismatch. We perform the analysis for the 50 U.S. states, using the HWOL data on online job ads coupled with CPS data on the unemployed to study misallocation of unemployed workers across geographical areas. Figure 13 shows the mismatch index $\mathcal{M}_t^h$ and the corresponding counterfactual experiment. We find that geographic mismatch is very low (about one tenth of the size of the index for 3-digit occupations, even though the number of active sectors is the same in both) and is essentially flat over the sample period under consideration. Unsurprisingly, the counterfactual unemployment computed based on the state-level index is essentially the same as the actual series, implying that geographical mismatch –across U.S. states– plays no role in the recent dynamics of U.S. unemployment.[28] This finding is consistent with other recent work that investigated the house-lock mechanism using different methods.[29]

[28]The JOLTS also provides limited geographic information, enabling us to study mismatch across the four broad Census regions. Our conclusions from the analysis of HWOL data on states are confirmed at this higher level of aggregation: we find no evidence of an increase in geographical mismatch.

[29]See, for example, Molloy, Smith, and Wozniak (2010) and Schulhofer-Wohl (2010). A related concern regarding geographic mobility was the observation that the rate of interstate migration in the U.S. reached a postwar low. How-

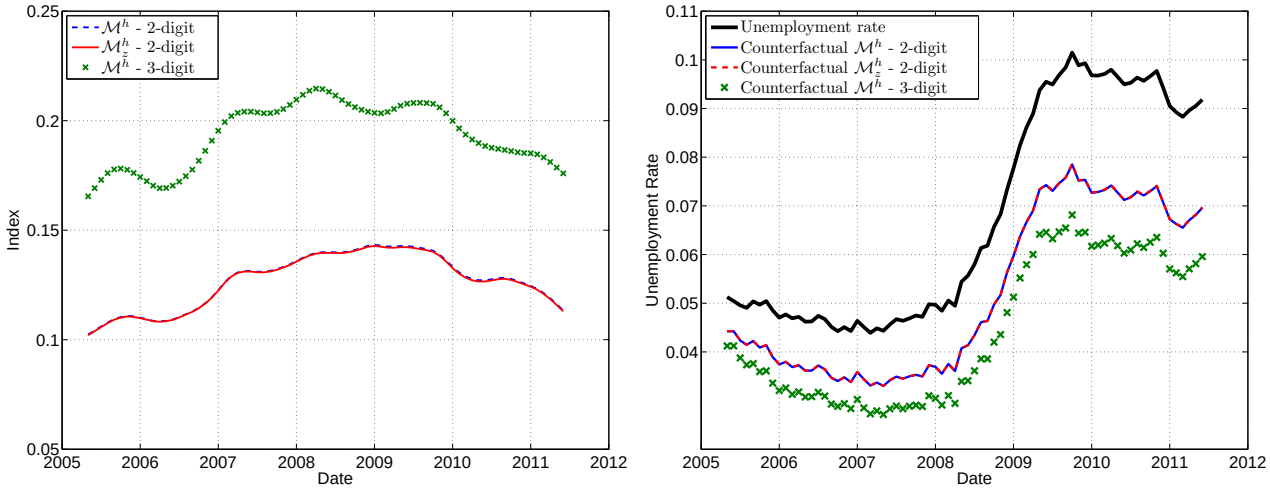

Figure 11: Mismatch index $\mathcal{M}^h_t$ by occupation (left panel) and corresponding counterfactuals (right panel).

5.4 Mismatch within education groups

Finally, we present our analysis of mismatch by education level, focusing on the following exercise. We define four education categories: less than high school diploma; high school diploma or equivalent; some college and Associate's degree; Bachelor's degree or higher. Then, we analyze mismatch by 2-digit occupation *within* these four education groups. This enables us to determine whether occupational mismatch has increased more or less for specific education categories.

The vacancy data for this analysis come from the HWOL series. As noted before, each ad recorded in HWOL constitutes an individual observation with a 6-digit occupation classification. We use this information, together with information from the BLS on the education content of 6-digit occupations, to construct vacancy counts for each 2-digit occupation by education group. In particular, the BLS provides information on the distribution of workers employed in each 6-digit occupation, broken down by their highest level of education attained.[30] We then allocate the 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.[31] Finally, we aggregate up to the 2-digit occupation level to obtain vacancy counts for each occupation by education cell.

Figures 14 and 15 illustrate our findings on occupational mismatch within each broad education

ever, 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 the effect of the change in procedures, they find that the annual interstate migration rate follows a smooth downward trend from 1996 to 2010.

[30]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.

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

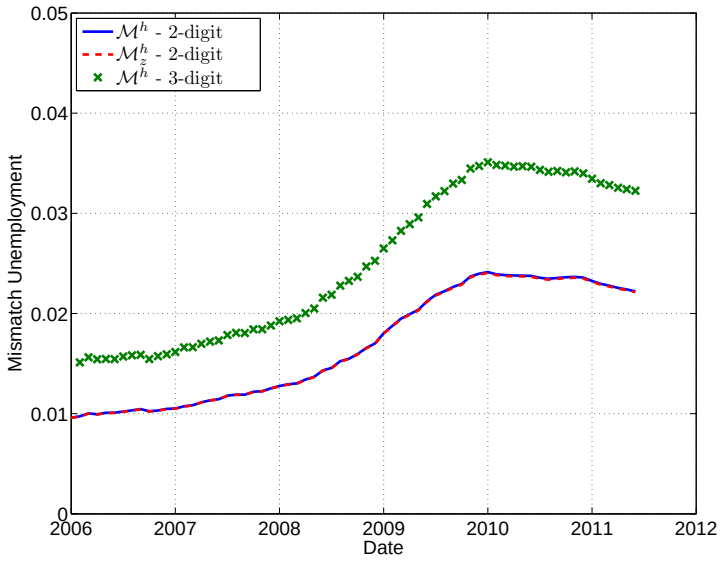

Figure 12: Mismatch unemployment measured at the occupation level.

category. The $\mathcal{M}_t^h$ mismatch indexes are shown in Figure 14 and the counterfactual unemployment exercises in Figure 15. Notice that actual unemployment varies considerably across the four panels in Figure 15, since we are plotting unemployment for workers within each educational attainment group. Unemployment dynamics differ greatly by education: for workers with less than high school, the unemployment rate rose from about 7% in 2006 to about 15% in 2010, an increase of about eight percentage points. The increase in unemployment rate over the same time period for high school graduates and those with some college was, respectively, 6 and 4.8 percentage points. For college graduates, the unemployment rate went from 2% to 4.7%, an increase of only 2.7 percentage points over the same period.

The occupational mismatch index rose within all four education groups, but more so in the some college and college categories. The counterfactual exercises 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 eight percentage point increase in unemployment for that group. For high school graduates, mismatch explains 1.2 (20%) out of the six percentage point increase in unemployment. For those with some college, mismatch explains about 1.4 (29%) out of a 4.8 percentage point rise in unemployment, and for college graduates 0.9 (33%) 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 third.

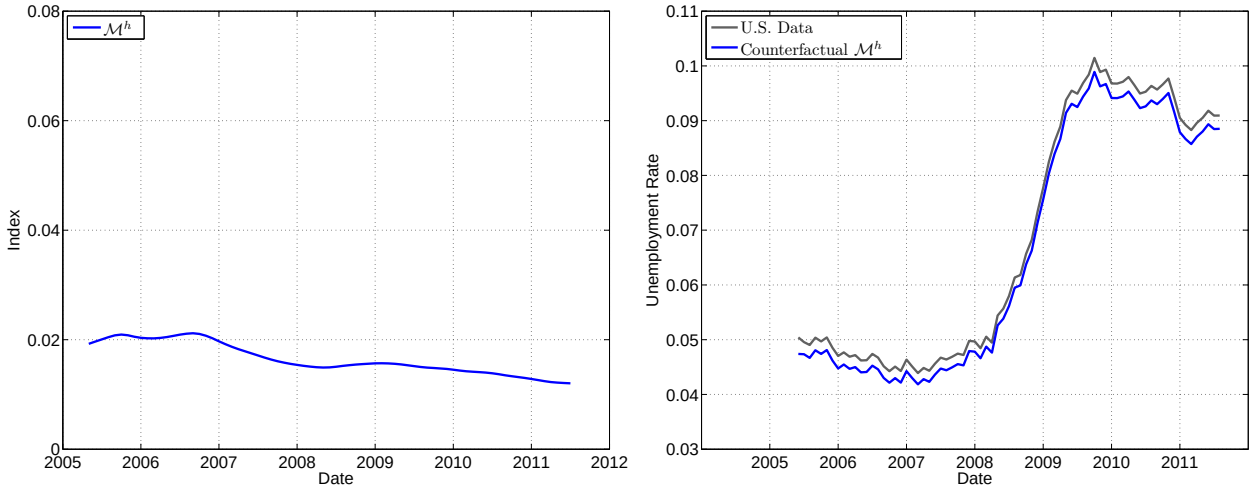

Figure 13: Mismatch index $\mathcal{M}_t^h$ by state (left panel) and corresponding counterfactual unemployment (right panel).

6 Conclusion

We have developed a theoretical framework that yields a meaningful notion of mismatch between unemployment and vacancies across separate labor markets (sectors) of an economy. Mismatch is defined as the distance between the empirically observed allocation and the allocation chosen by a planner who can freely move labor across markets. The solution to this planner's problem constitutes, in our view, a clean benchmark to measure the extent of misallocation of idle labor. The optimal allocation rule consists of an intuitive static condition that is easily generalizable to settings with heterogeneous productivities, match efficiencies, and job destruction rates across markets. These conditions are manipulated into mismatch indexes that capture the fraction of hires lost in the economy because of mismatch. We then use the resulting indexes to compute counterfactual series for unemployment in the absence of mismatch.

We exploit vacancy data by industry, occupation, geographic area, and education to compute our indexes for the period 2000-2010. We find that mismatch by industry or occupation can explain between 0.6 and 1.7 percentage points of the observed increase in the unemployment rate from the start of the recession to the end of 2009. The contribution of mismatch has declined somewhat since then. Our results indicate that the role of mismatch in explaining the increases in unemployment varies considerably by education. Occupational mismatch explains a substantial fraction of the rise in unemployment (one third) for high-educated workers while it is quantitatively less important for less-educated workers. Finally, we calculate geographic mismatch measures across Census regions and U.S. states and find no role for geographic mismatch in explaining the increase in the unemployment rate.

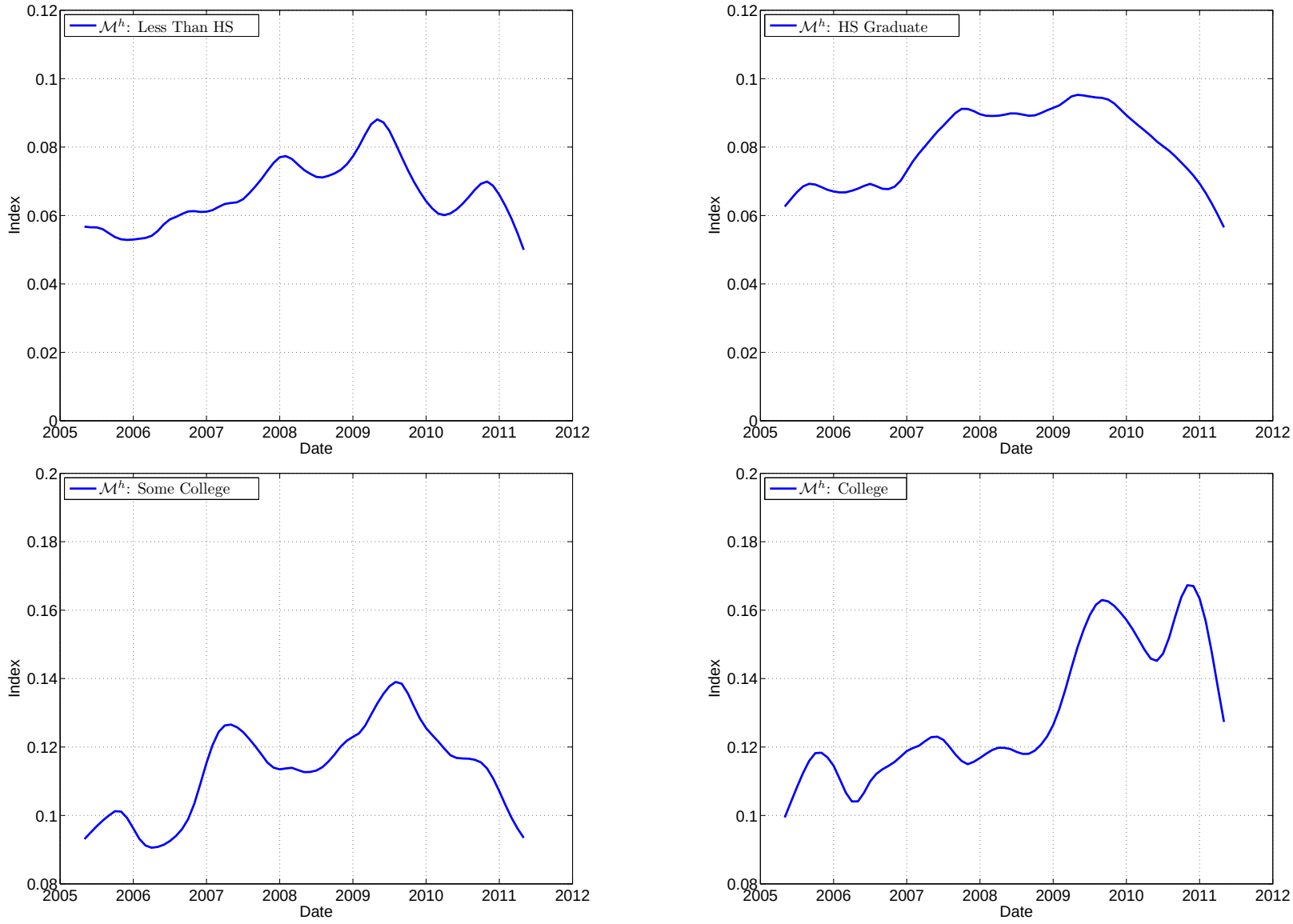

Figure 14: Mismatch index $\mathcal{M}_t^u$ by occupation within different education groups Top-left panel: Less than high school diploma. Top-right panel: High school diploma or equivalent. Bottom-left panel: Some college and Associate's degree. Bottom-right panel: Bachelor's degree or higher.

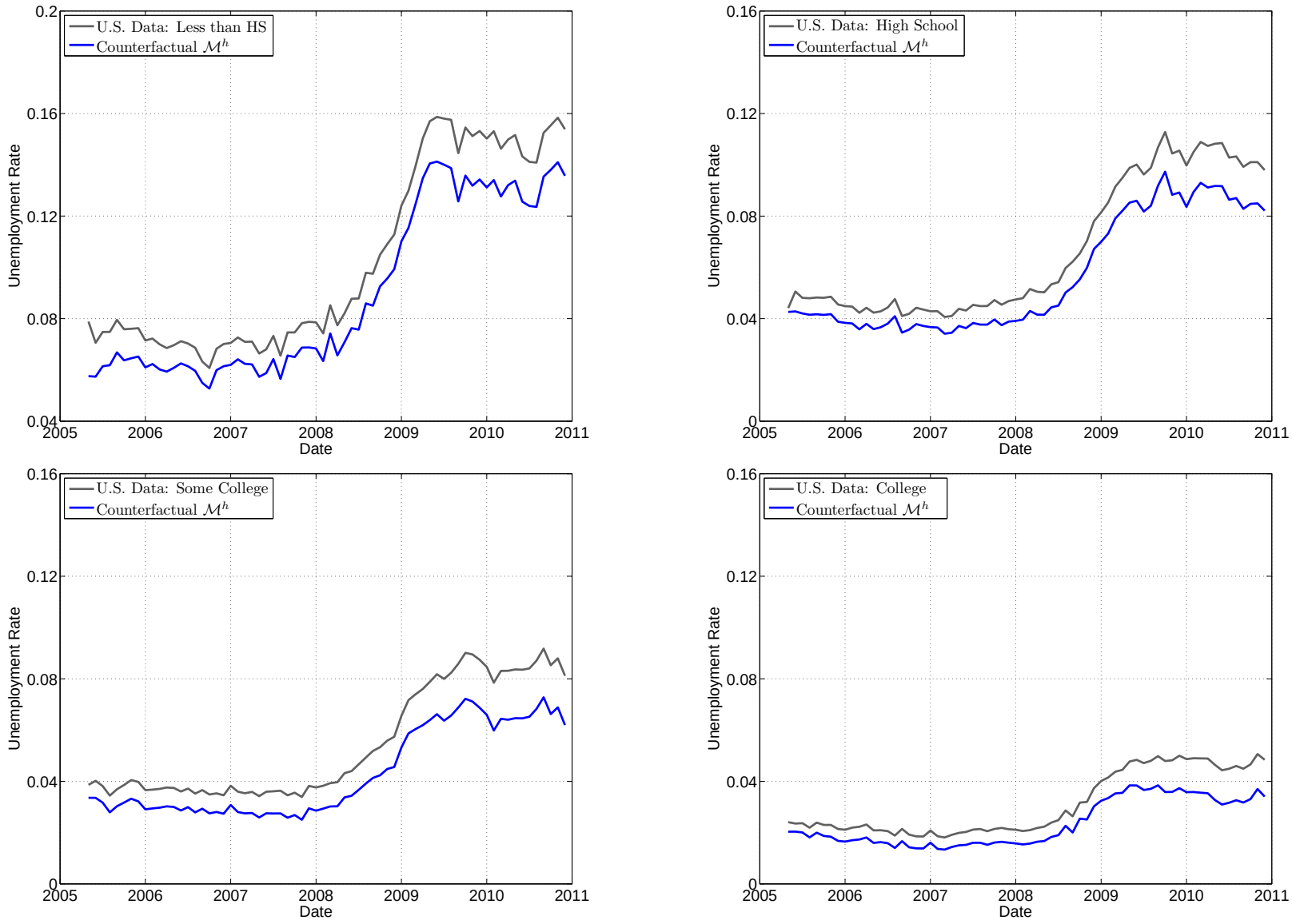

Figure 15: Counterfactual unemployment rate for different education groups. Top-left panel: Less than high school diploma. Top-right panel: High school diploma or equivalent. Bottom-left panel: Some college and Associate's degree. Bottom-right panel: Bachelor's degree or higher.

A1 Theoretical Appendix

This Appendix formally derives all the results discussed in the various generalizations of the baseline model in Section 2.3.

A1.1 Heterogenous productivities

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 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 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 planer's problem has two additional states: the current number of unemployed workers u, and the vector of productive efficiencies $\mathbf{z}$. The planner solves the problem:

$$
\begin{aligned}
V(u, \mathbf{e}; \mathbf{z}, \mathbf{v}, \phi, Z, \delta, \Phi) &= \max_{\{u_i, \ell'\}} \sum_{i=1}^{I} Z z_i (e_i + \gamma h_i) - \xi u + \beta \mathbb{E}\left[V(u', \mathbf{e}'; \mathbf{z}', \mathbf{v}', \phi', Z', \delta', \Phi')\right] \\
s.t. &: \\
\sum_{i=1}^{I} u_i &\leq u && \text{(A1)} \\
h_i &= \Phi \phi_i m(u_i, v_i) && \text{(A2)} \\
e_i' &= (1-\delta)(e_i + h_i) && \text{(A3)} \\
u' &= \ell' - \sum_{i=1}^{I} e_i' && \text{(A4)} \\
u_i &\in [0, u], \ell' \in [0, 1], && \text{(A5)} \\
&\Gamma_{Z,\delta,\Phi}(Z', \delta', \Phi'; Z, \delta, \Phi),\ \Gamma_{\mathbf{v}}(\mathbf{v}'; \mathbf{v}, Z', \delta', \Phi', \mathbf{z}'), \Gamma_{\phi}(\phi'; \phi), \Gamma_{\mathbf{z}}(\mathbf{z}'; \mathbf{z}) && \text{(A6)}
\end{aligned}
$$

The choice of how many unemployed workers u_i to allocate in the i market yields the first-order condition

$$
\gamma Z z_i \Phi \phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \beta \mathbb{E}\left[-V_u'(\cdot) + V_{e_i}'(\cdot)\right](1-\delta)\Phi \phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu, \qquad \text{(A7)}
$$

where μ is the multiplier on constraint $(A1)$. The Envelope conditions with respect to the states u and e_i yield:

$$
\begin{aligned}
V_u(u, \mathbf{e}; \mathbf{z}, \mathbf{v}, \phi, Z, \delta, \Phi) &= \mu - \xi && \text{(A8)} \\
V_{e_i}(u, \mathbf{e}; \mathbf{z}, \mathbf{v}, \phi, Z, \delta, \Phi) &= Z z_i + \beta(1-\delta)\mathbb{E}\left[V_{e_i}' - V_u'\right]. && \text{(A9)}
\end{aligned}
$$

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

$$\mathbb{E}\left[V_u\left(u', \mathbf{e}'; \mathbf{z}', \mathbf{v}', \phi', Z', \delta', \Phi'\right)\right] = 0, \tag{A10}$$

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 $(A10)$ with $(A8)$, 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.[32]

Consider now the Envelope condition $(A9)$ and make the additional assumption that z_i has a *linear* conditional mean function, i.e., $\mathbb{E}\left(z_i'\right) = \rho z_i$. We now conjecture that

$$V_{e_i}\left(u, \mathbf{e}; \mathbf{z}, \mathbf{v}, \phi, Z, \delta, \Phi\right) = z_i \Psi\left(Z, \delta, \Phi\right), \tag{A11}$$

where $\Psi\left(\cdot\right)$ is a function of Z, δ, and Φ alone. Using this conjecture into $(A9)$, we arrive at

$$V_{e_i}\left(u, \mathbf{e}; \mathbf{z}, \mathbf{v}, \phi, Z, \delta, \Phi\right) = Z z_i + \beta(1-\delta)\mathbb{E}\left[z_i' \Psi\left(Z', \delta', \Phi'\right)\right] = Z z_i + \beta(1-\delta)\rho z_i \mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right].$$

Or, using (A11):

$$\begin{aligned} z_i \Psi\left(Z, \delta, \Phi\right) &= Z z_i + \beta(1-\delta)\rho z_i \mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right] \\ \Psi\left(Z, \delta, \Phi\right) &= Z + \beta(1-\delta)\rho \mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right] \end{aligned}$$

which confirms the conjecture since $\mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right]$ is only a function of (Z, δ, Φ) by of the assumed structure for $\Gamma_{Z,\delta,\Phi}$.

Using this result into $(A7)$, together with $(A10)$, the optimality condition for the allocation of unemployed workers across sectors becomes

$$\gamma Z z_i \Phi \phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \beta\left(1-\delta\right)\rho \mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right] z_i \Phi \phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu, \tag{A12}$$

and rearranging:

$$z_i \phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \frac{\mu}{\gamma Z \Phi + \beta\left(1-\delta\right)\rho \Phi \mathbb{E}\left[\Psi\left(Z', \delta', \Phi'\right)\right]},$$

where the right hand side is a magnitude independent of i. We conclude that the optimal allocation rule equalizes the left hand side of this last equation across markets, yielding equation (8) in Section 2.3.1.

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

A1.2 Heterogenous sensitivities to aggregate shock

Let productivity in sector i be Z^{η_i} where η_i is a sector specific parameter measuring the elasticity to the aggregate shock Z. Let $\log Z$ follow a unit root process with innovation distributed as a $N(-\sigma_\varepsilon/2, \sigma_\varepsilon)$. Note that $\mathbb{E}\left[(Z')^{\eta_i}\right] = Z^{\eta_i} \exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)$. To simplify the exposition, set $\gamma = 1$ and assume that δ is constant over time. The envelope condition $(A9)$ becomes

$$V_{e_i} = Z^{\eta_i} + \beta(1-\delta)\mathbb{E}\left[V'_{e_i}\right].$$

which, solving forward and using the unit root assumption, yields

$$\mathbb{E}\left[V'_{e_i}\right] = \frac{Z^{\eta_i} \exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)}{1 - \beta(1-\delta)\exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)}.$$

Substituting the above expression for $\mathbb{E}\left[V'_{e_i}\right]$ into (the appropriately modified) equation $(A12)$, yields

$$Z^{\eta_i}\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) + \beta(1-\delta)\frac{Z^{\eta_i} \exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)}{1 - \beta(1-\delta)\exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)}\Phi\phi_i m_{u_i}\left(\frac{v_i}{u_i}\right) = \mu.$$

Rearranging, we conclude that the planner allocates unemployed workers so to equalize

$$\frac{Z^{\eta_i}}{1 - \beta(1-\delta)\exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right)}\phi_i m_{u_i}\left(\frac{v_i}{u_i^*}\right),$$

across sectors, which is expression (9) in Section $(2.3.2)$ in the main text. A necessary technical condition is $\beta(1-\delta)\exp\left(\eta_i(\eta_i - 1)\frac{\sigma_\varepsilon}{2}\right) < 1$.

A1.3 Heterogeneous destruction rates

We now relax the assumption that the destruction rate δ is common across sectors. Consider the environment of Section 2.3.1. Denote the idiosyncratic component of the exogenous destruction rate in sector i as δ_i. Then, the survival probability of a match is $(1-\delta)(1-\delta_i)$. To simplify the exposition, set $\gamma = 1$, and assume that $\{Z, \delta, z_i, \delta_i\}$ all follow independent unit root processes. The envelope condition $(A9)$ becomes

$$V_{e_i} = Zz_i + \beta(1-\delta)(1-\delta_i)\mathbb{E}\left[V'_{e_i}\right].$$

Solving forward, and using the unit root assumption, we arrive at:

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

which, substituted into (the appropriately modified) equation $(A12)$ yields

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

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

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

across sectors, which is expression (10) in Section $(2.3.3)$ in the main text.

A1.4 Endogenous separations

Consider the environment of Section $2.3.1$ 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

$$e_i' = \left(1-\delta\right)\left(e_i + h_i\right) - \sigma_i. \tag{A13}$$

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

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

$$\mathbb{E}\left[V_u\left(u', \mathbf{e}'; \mathbf{z}', \mathbf{v}', \phi', Z', \delta', \Phi'\right) - V_{e_i}\left(u', \mathbf{e}'; \mathbf{z}', \mathbf{v}', \phi', Z', \delta', \Phi'\right)\right] \begin{cases} < 0 & \rightarrow \sigma_i = 0 \\ = 0 & \rightarrow \sigma_i \in \left(0, \left(1-\delta\right)\left(e_i+h_i\right)\right) \\ > 0 & \rightarrow \sigma_i = \left(1-\delta\right)\left(e_i+h_i\right) \end{cases} \tag{A14}$$

depending on whether at the optimum a corner or interior solution arises. If the first-order condition $(A10)$ holds with equality, then the optimality condition $(A14)$ holds with the " $<$ " inequality and $\sigma_i = 0$. As a result, the planner's allocation rule (8) remains unchanged.

A2 Adjustment in sectoral unemployment count

Let u_{it} be the unemployed worker 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. Finally, 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 whose last job was in sector i. Let the total number of hires in sector j be h_t^j. Assume that the job-finding rate in sector j is the same for all unemployed, independently of the sector of provenance, with the possible exception if their previous job was in that same sector, in which case their job-finding rate is higher by a factor $\xi_t \geq 1$ than the average for the sector, or:

$$\frac{h_{it}^j}{u_{it}^j} = \xi_t \frac{h_t^j}{U_{jt}}, \text{ for all } i = 1, ..., I.$$

Rearranging the above equation and summing across all j yields, at every t, the I equations

$$u_{it} = \frac{1}{\xi_t} \sum_{j=1}^{I} \left(\frac{h_{it}^j}{h_{jt}} \right) U_{jt}, \text{ for all } i = 1, ..., I$$

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

$$\sum_{j=1}^{I} U_j = \sum_{j=1}^{I} u_j, \tag{A15}$$

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_i^j which accounted for less than 5% of hires h^j in any given sector. The estimated value of ξ is close to one.

A3 Measurement error in vacancies

Suppose that true vacancies (V_{it}) in market i are a factor $\mu_i^{\frac{1}{\alpha}}$ of the observed vacancies (v_{it}), i.e., $V_{it} = v_{it}\mu_i^{\frac{1}{\alpha}}$. For simplicity, consider the economy without heterogeneity in productive or matching efficiency of Section 2.1. The true mismatch index is

$$\begin{aligned} \mathcal{M}_{\mu t}^h &= 1 - \sum_{i=1}^{I} \left(\frac{V_{it}}{V_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} = 1 - \sum_{i=1}^{I} \left(\frac{v_{it}\mu_i^{\frac{1}{\alpha}}}{\sum_{i=1}^{I} v_{it}\mu_i^{\frac{1}{\alpha}}} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} \\ &= 1 - \sum_{i=1}^{I} \left(\frac{\mu_i}{\bar{\mu}_t} \right) \left(\frac{v_{it}}{v_t} \right)^{\alpha} \left(\frac{u_{it}}{u_t} \right)^{1-\alpha} \end{aligned} \tag{A16}$$

where $\bar{\mu}_t = \left[\sum_{i=1}^{I} \mu_i^{\frac{1}{\alpha}} \left(\frac{v_{it}}{v_t} \right) \right]^{\alpha}$. Note that the correction term $\mu_i/\bar{\mu}_t$ due to measurement error is exactly analogous to the correction term $\phi_i/\bar{\phi}_t$ for the index $\mathcal{M}_{\phi t}^h$ in (18). Is it possible to identify measurement error in vacancies μ_i in each sector? With a Cobb-Douglas specification, the true sectoral matching function is $h_{it} = \Phi_t V_{it}^{\alpha} u_{it}^{1-\alpha}$. Substituting observed variables measured with error in place of true ones, we arrive at

$$h_{it} = \Phi_t \cdot \mu_i \cdot v_{it}^{\alpha} u_{it}^{1-\alpha}.$$

Therefore, in a panel regression of log hires on log vacancies and log unemployment augmented with time dummies and fixed sector-specific effects, the estimated sector fixed-effect is precisely the measurement error in vacancies μ_i. Given an estimate of α, one can therefore obtain an estimate of μ_i in the same way we propose to estimate ϕ_i. To sum up, sectors where vacancies are especially underreported (i.e., $\mu_i >> 1$) will look like sectors with higher matching efficiency.

For the purpose of our measurement exercise, it therefore makes no difference whether we interpret ϕ_i as actual matching efficiency or measurement error, as long as we appropriately correct the mismatch index with the estimated fixed effects of the sectoral matching functions, as in (18).

References

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

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

[3] Barnichon, R. and A. Figura (2011). "What Drives Matching Efficiency? A Tale of Composition and Dispersion," FRB Staff WP 2011-10.

[4] Birchenall, J. A. (2010). "A Competitive Theory of Mismatch," mimeo, UCSB.

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

[6] Chari, V. V., P. Kehoe and E. McGrattan (2007). "Business Cycle Accounting," *Econometrica*, 75, 781836, 2007.

[7] Daly, M., B. Hobijn, A. Şahin and R. Valletta (2011). "A Rising Natural Rate of Unemployment: Transitory or Permanent?," mimeo.

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

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

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

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

[12] Hobijn, B. (2011). " Industry-Occupation Mix," mimeo.

[13] Hornstein, A., P. Krusell, and G.L. Violante (2007). "Technology-Policy Interaction in Frictional Labor Markets," *Review of Economic Studies*, 74, 1089-1124.

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

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

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

[17] 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.

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

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

[20] Ljungqvist, L. and T. J. Sargent (1998). "The European Unemployment Dilemma," *Journal of Political Economy* 106, 514-550.

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

[22] Molloy, R. C. Smith, and A. Wozniak (2010). "Internal Migration in the US: Updated Facts and Recent Trends," mimeo.

[23] Mortensen, D.T. (2009). "Island Matching," *Journal of Economic Theory*, 144(6), 2336-2353.

[24] Mortensen, D.T. and C.A. Pissarides (1999). "Unemployment Responses to Skill-biased Technology Shocks: the Role of Labour Market Policy," *Economic Journal*, 109(455), 242-265.

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

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

[27] Şahin, A., J. Song, G. Topa and G. Violante (2011). "Does Mismatch Explain the Recent Rise in U.K. Unemployment?," mimeo.

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

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

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

[31] Shimer R. (2007b). "Reassessing the Ins and Outs of Unemployment," mimeo.