Accounting for Mismatch Unemployment

Benedikt Herz and Thijs van Rens

January 2019

Appendices (for online publication)

A Model

This appendix provides the details on the derivations for the model. See Section 2.1 in the main text for a description of the model environment. Section A.1 below derives the efficient allocation, as discussed in Section 2.2 in the main text. Section A.2 derives the equilibrium, as used in Section 3 in the main text.

A.1 Efficient Allocation

A.1.1 Social Planner Problem

The social planner solves[29]

$$\max_{\{\{u_{it},v_{it}\}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \sum_i \Big(f(n_{it}; z_{it}) + b_{it} u_{it} - g(v_{it}; \kappa_{it}) + \lambda^u (1 - n_{it} - u_{it}) + \lambda^v (1 - n_{it} - v_{it}) \Big) \tag{33}$$

subject to

$$n_{it+1} = (1 - \delta_i) n_{it} + m(u_{it}, v_{it}; \phi_{it}) \tag{34}$$

Let $V(\{n_{it}\})$ denote the planner's value function in period t, which depends on the state variables n_{it} for each segment i. The planner's problem can be written in recursive form as the following Bellman equation,

$$V(\{n_{it}\}) = \max_{\{u_{it},v_{it}\}} \left\{ \sum_i \Big(f(n_{it}; z_{it}) + b_{it} u_{it} - g(v_{it}; \kappa_{it}) + \lambda^u (1 - n_{it} - u_{it}) + \lambda^v (1 - n_{it} - v_{it}) \Big) + \beta E_t V(\{n_{it+1}\}) \right\} \tag{35}$$

[29] Alternatively, we may drop the opportunity cost terms involving λ^u and λ^v, and instead impose the constraints that the planner takes the aggregate number of unemployed workers and vacancies as given.

$$\sum_i u_{it} = 1 - \sum_i n_{it} \tag{31}$$

$$\sum_i v_{it} = 1 - \sum_i n_{it} \tag{32}$$

In this case, (33) is the Lagrangian of the problem and λ_t^u and λ_t^v, which are then time varying, are the multipliers associated with constraints (31) and (32). We choose the opportunity cost formulation with time-invariant parameters λ^u and λ^v to facilitate the empirical implementation.

where n_{it+1} as in (34).

A.1.2 Efficiency Conditions

The first-order conditions for u_{it} and v_{it} are given by

$$m_u\left(u_{it}, v_{it}; \phi_{it}\right) S_{it} = \lambda^u - b_{it} \tag{36}$$

$$m_v\left(u_{it}, v_{it}; \phi_{it}\right) S_{it} = \lambda^v + g'\left(v_{it}; \kappa_{it}\right) \tag{37}$$

where $S_{it} = \beta E_t V_i\left(\{n_{it+1}\}\right)$ is the discounted expected value of having one more worker employed in segment i next period.

S_{it} is determined by the envelope condition for n_{it} (forwarding one period, taking conditional expectations and multiplying by β) and satisfies:

$$S_{it} = \beta E_t\left[f'\left(n_{it+1}; z_{it+1}\right) - \lambda^u - \lambda^v\right] + \beta\left(1-\delta_i\right) E_t S_{it+1} \tag{38}$$

Iterating forward,

$$S_{it} = \beta\sum_{s=0}^{\infty}\beta^s\left(1-\delta_i\right)^s E_t\left[f'\left(n_{it+s+1}; z_{it+s+1}\right) - \lambda^u - \lambda^v\right] = \frac{z_{it} - \lambda^u - \lambda^v}{r+\delta_i} \tag{39}$$

where the last equality follows if we further assume that $f\left(n_{it}; z_{it}\right) = z_{it}n_{it}$ is linear and z_{it} follows a random walk, so that $E_t f'\left(n_{it+s+1}; z_{it+s+1}\right) = z_{it}$.

A.1.3 Efficient allocation of unemployed workers and vacancies

Dividing the first-order condition for unemployment (36) by that of vacancies (37) and using a Cobb-Douglas matching function $m\left(u_{it}, v_{it}; \phi_{it}\right) = \phi_{it}u_{it}^{\mu}v_{it}^{1-\mu}$, we get an expression for the vacancy-unemployment ratio in each labor market segment.

$$\theta_{it} \equiv \frac{v_{it}}{u_{it}} = \frac{1-\mu}{\mu}\frac{\lambda^u - b_{it}}{\lambda^v + g'\left(v_{it}; \kappa_{it}\right)} \tag{40}$$

Substituting condition (40) into the expression for the job-finding probability, $p_{it} = \phi_{it}\theta_{it}^{1-\mu}$, gives equation (1) in the main text.

Condition (40) is the same efficiency condition as in Şahin, Song, Topa, and Violante (2014) if we set $g\left(v_{it}; \kappa_{it}\right) = \frac{1}{1+\varepsilon}\kappa_{it}^{\varepsilon}v_{it}^{1+\varepsilon}$ and $\lambda^v = 0$ (free entry of vacancies). Substituting these assumptions into (40), substituting the result into (36), and solving for v_{it} gives,

$$v_{it} = \frac{1}{\kappa_{it}}\left(\frac{1-\mu}{\mu}\right)^{1/\varepsilon}\left(\frac{1}{\lambda^u - b_{it}}\right)^{-\frac{\mu/\varepsilon}{1-\mu}}\left(\mu\phi_{it}S_{it}\right)^{\frac{1/\varepsilon}{1-\mu}} \tag{41}$$

which is equation (A36) in their paper if we set $\mu = 1-\alpha$, $\lambda^u = \tilde{\mu}$, $b_{it} = 0$, $\phi_{it} = \Phi\phi_i$ and $S_{it} = \frac{z_{it}-\lambda^u-\lambda^v}{r+\delta_i} = \frac{Zz_i}{1-\beta(1-\Delta)(1-\delta_i)}$ to be consistent with their notation and assumptions (note that Zz_i in Şahin et al. is defined net of the output of the nonemployed, i.e. equal to $z_{it} - \lambda^u$ in our notation).

In order to compare to the baseline in Şahin, Song, Topa, and Violante (2014), in which the vacancy distribution is exogenous, we simply drop the first-order condition for vacancies (37), so that the efficient allocation is described by conditions (36) and (39), which is condition (2) in the main text.

A.1.4 Productivity and matching efficiency

If we assume $b_{it} = b_t$ and $g'(v_{it}; \kappa_{it}) = \kappa_t$ in addition to a Cobb-Douglas form for the matching function, and substitute (40) back into (36), we get an efficiency condition that does not depend on the distributions of unemployed worker and vacancies.

$$\phi_{it} S_{it} = \tfrac{1}{\mu} \left(\tfrac{\mu}{1-\mu}\right)^{1-\mu} (\lambda^u - b_t)^{\mu} (\lambda^v + \kappa_t)^{1-\mu} \tag{42}$$

In the efficient allocation, $S_{it} = E_t [f'(n_{it+s+1}; z_{it+s+1}) - \lambda^u - \lambda^v] / (r + \delta_i)$ must be inversely proportional to matching efficiency across labor market segments, which implies that S_{it} must be equalized if matching efficiency is constant across labor market segments, as discussed in Section 2.3 in the main text.

A.2 Equilibrium Allocation

There are two types of agents in our economy: a large representative household, consisting of a measure 1 of workers, who may be employed, unemployed or not in the labor force, and a large representative firm, consisting of a measure 1 of positions, which may be filled, vacant or closed.

A.2.1 Households

The household does not have a technology for intertemporal consumption smoothing nor a motive to do so (the utility function is linear in consumption), so that maximizing utility is equivalent to maximizing consumption and maximizing income. Thus, the household chooses how many unemployed workers to allocate to each labor market segment in order to solve

$$\max_{\{\{u_{it}\}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \sum_i \Big(w_{it} n_{it} + b_{it} u_{it} + \lambda^u (1 - n_{it} - u_{it}) \Big) \tag{43}$$

subject to

$$n_{it+1} = (1 - \delta_i) n_{it} + p_{it} u_{it} \tag{44}$$

and taking w_{it} (wages) and p_{it} (job finding probabilities) as given. The endogenous variables u_{it} (number of unemployed workers) and n_{it} (employment), the exogenous variable b_{it} (home production of the unemployed), and the parameters δ_i (separation probabilities), β (discount factor) and λ^u (home production of workers not participating in the labor force) are the same as for the social planner problem, and were introduced in the main text.

Writing the problem in recursive form, with $\{n_{it}\}$ as the endogenous state variables, we get the following first-order condition for u_{it},

$$p_{it}S_{it}^{W} = \lambda^{u} - b_{it} \tag{45}$$

which is worker mobility condition (3) in the main text.

S_{it}^{W} is the discounted expected value to the household of having one more worker employed in segment i next period, and is found from the envelope condition for n_{it}, by forwarding one period, taking conditional expectations and multiplying by β,

$$S_{it}^{W} = \beta E_{t}\left[w_{it+1} - \lambda^{u}\right] + \beta\left(1-\delta_{i}\right)E_{t}S_{it+1}^{W} = \beta\sum_{s=0}^{\infty}\beta^{s}\left(1-\delta_{i}\right)^{s}E_{t}\left[w_{it+s+1} - \lambda^{u}\right] \tag{46}$$

where the last equality follows from iterating forward. If we further assume that wages follow a random walk, then $S_{it}^{W} = \left(w_{it} - \lambda^{u}\right)/\left(r+\delta_{i}\right)$.

A.2.2 Firms

The firm chooses how many vacancies to post in each labor market segment. To make the problem analogous to the household's problem, we assume that there is an opportunity cost of having a filled or unfilled position equal to λ^{v}, which includes free entry of vacancies as a special case by setting $\lambda^{v} = 0$. Since the utility function of the household is linear, the firm uses the same discount factor $\beta = 1/\left(1+r\right)$ as the household. Thus, the firm solves

$$\max_{\{\{v_{it}\}\}_{t=0}^{\infty}} E_{0}\sum_{t=0}^{\infty}\beta^{t}\sum_{i}\Big(f\left(n_{it};z_{it}\right) - w_{it}n_{it} - g\left(v_{it};\kappa_{it}\right) + \lambda^{v}\left(1-n_{it}-v_{it}\right)\Big) \tag{47}$$

subject to

$$n_{it+1} = \left(1-\delta_{i}\right)n_{it} + q_{it}v_{it} \tag{48}$$

and taking w_{it} (wages) and q_{it} (job filling probabilities) as given. The endogenous variables v_{it} (number of vacancies) and n_{it} (employment), the exogenous variables z_{it} (production efficiency) and κ_{it} (vacancy cost parameter), and the parameters δ_{i} (separation probabilities), β (discount factor) and λ^{v} (opportunity cost of keeping a position open) are the same as for the social planner problem, and were introduced in the main text.

The first-order condition for v_{it},

$$q_{it}S_{it}^{F} = \lambda^{v} + g'\left(v_{it};\kappa_{it}\right) \tag{49}$$

is job mobility condition (4) in the main text, with $\lambda^{v} = 0$ if we assume free entry of vacancies, and where S_{it}^{F} is the discounted expected value to the firm of having one more worker employed in segment i next period.

$$\begin{aligned} S_{it}^F &= \beta E_t \left[f\left(n_{it+1}; z_{it+1}\right) - w_{it+1} - \lambda^v \right] + \beta \left(1 - \delta_i\right) E_t S_{it+1}^F \\ &= \beta \sum_{s=0}^{\infty} \beta^s \left(1 - \delta_i\right)^s E_t \left[f\left(n_{it+s+1}; z_{it+s+1}\right) - w_{it+s+1} - \lambda^v \right] \end{aligned} \tag{50}$$

This expression for firm surplus simplifies to $S_{it}^F = \left(f'\left(n_{it}; z_{it}\right) - w_{it} - \lambda^v\right) / \left(r + \delta_i\right)$ if we further assume that profits follow a random walk.

B Effect of Mismatch on the Aggregate Job-Finding Rate

The aggregate job-finding probability is given by the average job-finding probability across labor market segments, weighted by the number of unemployed workers in each segment.

$$\bar{p}_t = \frac{\sum_i u_{it} p_{it}}{\sum_i u_{it}} = \phi_t \frac{\sum_i u_{it} \theta_{it}^{1-\mu}}{\sum_i u_{it}} \tag{51}$$

where $\theta_{it} = v_{it}/u_{it}$ is the vacancy-unemployment ratio in segment i at time t.

Reallocating unemployed workers and vacancies, keeping constant the total number of each, does not affect the aggregate vacancy-unemployment ratio. To see this, let u_{it}^*, v_{it}^* and θ_{it}^* denote the allocation of unemployed workers and vacancies chosen by the planner, and the resulting allocation of the vacancy-unemployment ratio. Then, the aggregate vacancy-unemployment ratio in the planner allocation,

$$\frac{\sum_i u_{it}^* \theta_{it}^*}{\sum_i u_{it}^*} = \frac{\sum_i v_{it}^*}{\sum_i u_{it}^*} = \frac{\sum_i v_{it}}{\sum_i u_{it}} = \frac{\sum_i u_{it} \theta_{it}}{\sum_i u_{it}} \tag{52}$$

equals the aggregate vacancy-unemployment ratio before reallocation. Therefore, in the planner allocation the vacancy-unemployment ratio in each labor market segment equals the aggregate ratio, so that the efficient job-finding rate is given by

$$\bar{p}_t^* = \phi_t \left(\frac{\sum_i u_{it} \theta_{it}}{\sum_i u_{it}} \right)^{1-\mu} = \left(\frac{\sum_i u_{it} p_{it}^{\frac{1}{1-\mu}}}{\sum_i u_{it}} \right)^{1-\mu} \tag{53}$$

To make explicit the aggregate effect of dispersion in θ_{it} and p_{it}, we approximate the aggregate job-finding rate $\bar{p}_t$ as in equation (51) and the job-finding rate without mismatch $\bar{p}_t^*$ as in equation (53) by assuming the distribution of p_{it} is log-normal. If p_{it} is log-normally distributed, i.e., $\hat{p}_{it}$ has a normal distribution, then

$$\log E\left[p_{it}\right] = E\left[\hat{p}_{it}\right] + \tfrac{1}{2} V\left[\hat{p}_{it}\right] \tag{54}$$

where E and V are the unemployment-weighted cross-sectional expectation and variance operators. In addition, $p_{it}^{\frac{1}{1-\mu}}$ must be log-normal as well, so that

$$\log E\left[p_{it}^{\frac{1}{1-\mu}}\right]=\frac{1}{1-\mu}E\left[\hat{p}_{it}\right]+\tfrac{1}{2}\left(\frac{1}{1-\mu}\right)^{2}V\left[\hat{p}_{it}\right] \tag{55}$$

Substituting this into equations (51) for $\bar{p}_t$ and (53) for $\bar{p}_t^*$ gives,

$$\log\bar{p}_t^*-\log\bar{p}_t=(1-\mu)\left(\frac{1}{1-\mu}E\left[\hat{p}_{it}\right]+\tfrac{1}{2}\left(\frac{1}{1-\mu}\right)^{2}V\left[\hat{p}_{it}\right]\right)-E\left[\hat{p}_{it}\right]-\tfrac{1}{2}V\left[\hat{p}_{it}\right]=\tfrac{1}{2}\frac{\mu}{1-\mu}V\left[\hat{p}_{it}\right] \tag{56}$$

which is expression (23) in the main text.

C Counterfactual Decompositions

Equation (23) in the main text expresses the relative contribution of mismatch to the aggregate job finding rate in terms of the deviations from the four no-mismatch equilibrium conditions.

$$\log\bar{p}_t^*-\log\bar{p}_t=\tfrac{1}{2}\mu\left(1-\mu\right)V\left[\hat{\gamma}_{it}^{WM}-\hat{\gamma}_{it}^{JM}-\hat{\gamma}_{it}^{WD}\right] \tag{57}$$

We use this expression for counterfactual analysis, where we ‘shut down’ a friction by setting the corresponding wedge equal to zero, e.g. to evaluate the job finding rate in the absence of worker mobility frictions we set $\hat{\gamma}_{it}^{WM}=0$.

There are two ways to define the contribution of a particular friction to unemployment. First, we can shut down the friction, leaving all other frictions in place, and compare the resulting counterfactual aggregate job finding rate to the actual job finding rate.

$$\Delta\log\bar{p}_t^{WM,1}=\tfrac{1}{2}\mu\left(1-\mu\right)\left(V\left[\hat{\gamma}_{it}^{WM}-\hat{\gamma}_{it}^{JM}-\hat{\gamma}_{it}^{WD}\right]-V\left[0-\hat{\gamma}_{it}^{JM}-\hat{\gamma}_{it}^{WD}\right]\right) \tag{58}$$

Alternatively, we can shut down all other frictions, leaving only the friction we are considering in place, and compare the resulting counterfactual job finding rate to the job finding rate that would prevail in the absence of all sources of mismatch.

$$\Delta\log\bar{p}_t^{WM,2}=\tfrac{1}{2}\mu\left(1-\mu\right)\left(V\left[\hat{\gamma}_{it}^{WM}\right]-0\right) \tag{59}$$

The difference between the two estimators is that $\Delta\log\bar{p}_t^{WM,1}$ includes the covariance terms of $\hat{\gamma}_{it}^{WM}$ with the other wedges, whereas $\Delta\log\bar{p}_t^{WM,2}$ does not. The contribution of all frictions adds up to more than the total amount of mismatch by the first estimator, and to less than the total by the second estimator.

By combining both estimators, we can disentangle the direct contribution of a friction from its contribution through its correlation with other frictions and thus design an additive decomposition:

$$\Delta \log \bar{p}_t^{WM,1} = \tfrac{1}{2}\mu\left(1-\mu\right)\left(V\left[\hat{\gamma}_{it}^{WM}\right] - 2Cov\left[\hat{\gamma}_{it}^{WM}, \hat{\gamma}_{it}^{JM}\right] - 2Cov\left[\hat{\gamma}_{it}^{WM}, \hat{\gamma}_{it}^{WD}\right]\right) \tag{60}$$

so that

$$\begin{aligned}\Delta \log \bar{p}_t^{WM} &= \tfrac{1}{2}\left(\Delta \log \bar{p}_t^{WM,1} + \Delta \log \bar{p}_t^{WM,2}\right) \\ &= V\left[\hat{\gamma}_{it}^{WM}\right] - 2Cov\left[\hat{\gamma}_{it}^{WM}, \hat{\gamma}_{it}^{JM}\right] - 2Cov\left[\hat{\gamma}_{it}^{WM}, \hat{\gamma}_{it}^{WD}\right] \end{aligned} \tag{61}$$

and similarly for the other frictions. Because this estimator includes half of the covariance terms of $\hat{\gamma}_{it}^{WM}$ with the other wedges, with the remaining half being attributed to the other frictions, it satisfies

$$\Delta \log \bar{p}_t^{WM} + \Delta \log \bar{p}_t^{JM} + \Delta \log \bar{p}_t^{WD} = \log \bar{p}_t^{*} - \log \bar{p}_t \tag{62}$$

The contribution of all frictions adds up to overall mismatch.

D Match Surplus with Time-Varying Payoffs and Turnover

In order to be able to solve forward for match surplus, take a linear approximation of the Bellman equation around $\delta_{it} = \delta_i^*$ and $S_{it} = S_i^*$.

$$\left(1+r\right)S_{it} = y_{it} + E_t\left[\left(1-\delta_{it+1}\right)S_{it+1}\right] \simeq y_{it} + \left(1-\delta_i^*\right)E_t S_{it+1} + E_t\left[\delta_i^* - \delta_{it+1}\right]S_i^* \tag{63}$$

Now, we can solve forward as if the separation probability were constant:

$$\begin{aligned} S_{it} &\simeq \frac{1}{1+r}\left\{y_{it} + E_t\left[\delta_i^* - \delta_{it+1}\right]S_i^*\right\} + \frac{1-\delta_i^*}{1+r}E_t S_{it+1} \\ &= \frac{1}{1+r}\sum_{s=0}^{\infty}\left(\frac{1-\delta_i^*}{1+r}\right)^s E_t\left[y_{it+s} + \left(\delta_i^* - \delta_{it+s+1}\right)S_i^*\right] \end{aligned} \tag{64}$$

From the autoregressive processes for payoffs and separation rates,

$$y_{it+1}^k = \left(1-\rho_y^k\right)y_{it}^k + \rho_y^k \bar{y}_t^k + \varepsilon_{y,it+1}^k \Rightarrow E_t y_{it+s}^k = \bar{y}_t^k + \left(1-\rho_y^k\right)^s\left(y_{it}^k - \bar{y}_t^k\right) \tag{65}$$

$$\delta_{it+1} = \left(1-\rho_\delta\right)\delta_{it} + \rho_\delta\bar{\delta}_t + \varepsilon_{\tau,it+1}^k \Rightarrow E_t\delta_{it+s} = \bar{\delta}_t + \left(1-\rho_\delta\right)^s\left(\delta_{it} - \bar{\delta}_t\right) \tag{66}$$

we get (dropping the k superscripts for simplicity)

$$E_t y_{it+s} = \bar{y}_t + \left(1-\rho_y\right)^s\left(y_{it} - \bar{y}_t\right) \tag{67}$$

$$E_t\left[\delta_i^* - \delta_{it+s+1}\right] = \delta_i^* - \bar{\delta}_t + \left(1-\rho_\delta\right)^{s+1}\left(\bar{\delta}_t - \delta_{it}\right) \tag{68}$$

Substituting into the expression for surplus

$$
\begin{aligned}
S_{it} &\simeq \frac{1}{1+r}\sum_{s=0}^{\infty}\left(\frac{1-\delta_i^*}{1+r}\right)^s\left\{\bar{y}_t+(1-\rho_y)^s(y_{it}-\bar{y}_t)+\left(\delta_i^*-\bar{\delta}_t\right)S_i^*+(1-\rho_\delta)^{s+1}(\bar{\tau}_t-\tau_{it})S_i^*\right\}\\
&=\frac{1}{1+r}\sum_{s=0}^{\infty}\left(\frac{1-\delta_i^*}{1+r}\right)^s\left\{\bar{y}_t+\left(\delta_i^*-\bar{\delta}_t\right)S_i^*\right\}+\frac{1}{1+r}\sum_{s=0}^{\infty}\left(\frac{(1-\delta_i^*)(1-\rho_y)}{1+r}\right)^s(y_{it}-\bar{y}_t)\\
&+\frac{1-\rho_\delta}{1+r}\sum_{s=0}^{\infty}\left(\frac{(1-\delta_i^*)(1-\rho_\delta)}{1+r}\right)^s\left(\bar{\delta}_t-\delta_{it}\right)S_i^*\\
&=\frac{\bar{y}_t+\left(\delta_i^*-\bar{\delta}_t\right)S_i^*}{r+\delta_i^*}+\frac{y_{it}-\bar{y}_t}{r+\delta_i^*+\rho_y-\rho_y\delta_i^*}+\frac{(1-\rho_\tau)\left(\bar{\delta}_t-\delta_{it}\right)S_i^*}{r+\delta_i^*+\rho_\delta-\rho_\delta\delta_i^*}\\
&\simeq\frac{\bar{y}_t+\left(\delta_i^*-\bar{\delta}_t\right)S_i^*}{r+\delta_i^*}+\frac{y_{it}-\bar{y}_t}{r+\delta_i^*+\rho_y}+\frac{(1-\rho_\delta)\left(\bar{\delta}_t-\delta_{it}\right)S_i^*}{r+\delta_i^*+\rho_\delta}
\end{aligned}
\tag{69}
$$

Finally, setting $\delta_i^* = \delta_{it}$ and $S_i^* = S_{it}$ and rearranging we get the expression in the main text.

$$
S_{it} \simeq \frac{(r+\delta_{it})(r+\delta_{it}+\rho_\delta)}{(r+\delta_{it})(r+\delta_{it}+\rho_\delta)+\rho_\delta(1+r+\delta_{it})\left(\bar{\delta}_t-\delta_{it}\right)}\left(\frac{\bar{y}_t}{r+\delta_{it}}+\frac{y_{it}-\bar{y}_t}{r+\delta_{it}+\rho_y}\right) \tag{70}
$$

E Disaggregation and the Level of Mismatch

Unemployment due to mismatch across states and 2-digit industries is around one order of magnitude smaller than unemployment due to mismatch across 3-digit occupations. As mentioned in the main text, we believe this is because states and 2-digit industries are not sufficiently disaggregated, and most of the mismatch is within a state or an industry. In this appendix, we provide some suggestive evidence for this claim.

We address the aggregation issue in two ways. First, we disaggregate further. For the purposes of this appendix only, we use data that are disaggregated by both state and industry. Instead of 50 states or 33 industries, this gives us $50 * 33 = 1650$ labor market segments. Although 1650 submarkets is probably a more realistic segmentation of the US labor market, it is in all likelihood still to coarse. Therefore, the second part of our solution is to find a correction factor that relates the observed amount of mismatch in our data to the amount of mismatch we would observe if we were to disaggregate to the right level, based on earlier work by Barnichon and Figura (2015).[30]

Disaggregation by both states and industries, while alleviating the aggregation problem, gives rise to a different bias because of sampling error. The data we use are survey-based and consequently we have only about $23,000$ unemployed workers per year, which means that the 1650 labor market segments on average contain only 14 observations and because not all states and industries are equally large, some cells are even much smaller than that. As a result, our estimates for the job finding rate in each segment will be

[30] The exercise we describe here is not present in the published article, but may be found in the December 2011 working paper version, available at https://www.bde.es/f/webbde/GAP/Secciones/SalaPrensa/Agenda/Eventos/12/May/barnichon_figura.pdf.

very imprecise. This sampling error will translate into dispersion across segments and bias our estimate for the amount of mismatch unemployment. We address this issue by estimating the variance of the sampling error in each segment and correcting the estimated variance of the job finding rates by subtracting the average variance of the sampling error.[31]

E.1 Correction Factor

An ideal labor market segment would consist of very similar jobs within a geographic area that allows workers to commute to these jobs without moving house. Using UK data, Barnichon and Figura (2015) estimate the correct level of disaggregation would be to use 232 so-called travel-to-work areas and 353 detailed occupational groups. They then aggregate these data to a level that is comparable to US states and major occupational categories and find that the observed amount of mismatch decreases by a factor 6. We argue that a similar correction factor is appropriate for our estimate of mismatch across 1650 state-industry segments.

From equation (23), we know that mismatch is approximately proportional to the variance of log job-finding rates, $V\left[\hat{p}_{it}\right]$. Barnichon and Figura show that

$$\ln\left(V_n\left[\hat{p}_i\right]\right) \simeq \ln a_0 + a_{geo}\ln n_{geo} + a_{occ}\ln n_{occ} \tag{71}$$

where V_n is the variance of $\hat{p}_i$ based on a higher level of aggregation and $n = N/N^{CF}$ is the ratio of the observed versus the correct number of labor market segments. They also estimate the parameters of this relation using UK data to and find $a_{geo} = 0.13$ and $a_{occ} = 0.67$. This implies

$$\ln\left(\frac{V\left[\hat{p}_i^{CF}\right]}{V\left[\hat{p}_i\right]}\right) = a_{geo}\ln\left(\frac{1}{n_{geo}}\right) + a_{occ}\ln\left(\frac{1}{n_{occ}}\right) \tag{72}$$

because by assumption $\hat{p}_i^{CF}$ are the finding rates for the right level of disaggregation so that $n_{geo}^{CF} = n_{occ}^{CF} = 1$.

In the UK data that Barnichon and Figura use, the correct number of geographic areas is about 232 (travel to work areas). The US population is larger than the UK population, but the land area is larger as well. Therefore, Barnichon and Figura assume the number of geographic units is the same in the same in the two countries. Since we work with 50 states, $1/n_{geo} = 232/50 = 4.64$. The same UK data have 353 detailed occupational groups, which should be similar in the US. We use 33 broad industries. Assuming these broad industry categories are comparable to broad occupations cate-

[31] Workers in each segment find a job with probability p_i. The variance of the realization of this Bernoulli process equals $p_i\left(1-p_i\right)$, so that the variance of the observed mean probability is equal to $p_i\left(1-p_i\right)/N_i$, where N_i is the number of observations in segment i. The variance of the signal in p_i across segments, by the ANOVA formula, is then given by the observed variance $var\left(p_i\right)$ minus the average variance of the sampling error $E\left[p_i\left(1-p_i\right)/N_i\right]$. We do not use segments with less than 5 observations because these would contribute more noise than signal.

gories, we get $1/n_{occ} = 353/33 = 10.7$. This implies a correction factor for the variance of labor market tightness of,

$$\frac{V\left[\hat{p}_i^{CF}\right]}{V\left[\hat{p}_i^{geo*ind}\right]} = \exp\left(0.13\ln\left(4.64\right) + 0.67\ln\left(10.7\right)\right) = 6.0 \tag{73}$$

which is the same correction factor that Barnichon and Figura used.

E.2 Results

Mismatch across state*industry segments contributes 15% to unemployment, comparable to mismatch across occupation-state segments and substantially more than mismatch across states or industries only. The bias because of sampling error is fairly small, bringing the contribution of mismatch down to 14%, indicating the dispersion in job-finding rates across segments is large compared to the sampling error. After muliplying by 6 to correct for aggregation, these estimates suggest that mismatch is responsible for 84% of unemployment. It is important to note that a good amount of guesswork was needed for the aggregation correction and the estimate is therefore rather imprecise. Nevertheless, these estimates indicate that it is quite possible that mismatch across states and industries is of the same order of magnitude as mismatch across detailed occupations, and that mismatch is an important contributor to unemployment.

F Additional Tables

Table 2A
State-level data, cell sizes 1979-2015

		job finding rate			wage		
		min	mean	max	min	mean	max
Alabama	AL	292	630	12656	1484	2020	2808
Alaska	AK	409	780	11703	1600	2005	2400
Arizona	AZ	323	528	12178	1584	1943	2650
Arkansas	AR	322	543	11940	1480	1867	2433
California	CA	2079	4069	67582	8996	12626	14951
Colorado	CO	296	701	16924	1776	2777	3637
Connecticut	CT	159	602	16889	1590	2514	3781
Delaware	DE	210	441	11684	1078	1974	2706
District of Columbia	DC	196	549	10318	671	1542	2380
Florida	FL	960	1663	35367	4633	6318	8201
Georgia	GA	374	721	16769	1999	2780	3846
Hawaii	HI	179	390	10919	1301	1862	2375
Idaho	ID	252	569	11443	1522	1939	2321
Illinois	IL	1002	1839	33016	4506	6317	7992
Indiana	IN	256	737	17025	2100	2626	3853
Iowa	IA	218	576	15461	2011	2651	3438
Kansas	KS	266	497	13196	1950	2287	2922
Kentucky	KY	338	632	13105	1785	2070	2779
Louisiana	LA	259	571	13233	1310	1758	2983
Maine	ME	254	593	14862	1418	2176	3200
Maryland	MD	239	613	16907	1647	2608	3839
Massachusetts	MA	400	1028	33071	2330	4532	7689
Michigan	MI	707	1831	32804	3258	5498	7984
Minnesota	MN	274	695	19757	2001	3076	4367
Mississippi	MS	297	584	12266	1159	1773	2633
Missouri	MO	253	674	14410	1870	2449	3000
Montana	MT	265	546	12356	1265	1816	2518
Nebraska	NE	193	393	13611	1483	2396	2977
Nevada	NV	359	659	15305	1608	2244	3456
New Hampshire	NH	177	472	17533	1429	2416	3875
New Jersey	NJ	613	1291	33084	3055	5123	7987
New Mexico	NM	249	514	11069	1050	1662	2330
New York	NY	1282	2401	53144	5950	8921	12941
North Carolina	NC	488	1052	35075	2759	4638	8276
North Dakota	ND	240	400	13214	1644	2154	2466
Ohio	OH	767	1740	35137	3964	6143	8497
Oklahoma	OK	214	486	12383	1414	1971	2538
Oregon	OR	398	713	13963	1539	2010	2933
Pennsylvania	PA	888	1700	33274	4347	6327	8188
Rhode Island	RI	219	610	13968	1142	2092	3170
South Carolina	SC	238	564	11190	1607	2017	2730
South Dakota	SD	233	440	13238	1733	2346	2825
Tennessee	TN	324	603	11655	1912	2121	2523
Texas	TX	1316	2090	38910	6873	7870	8576
Utah	UT	249	480	14645	1798	2157	3158
Vermont	VT	184	424	12303	1289	1931	2611
Virginia	VA	216	598	15987	2338	2931	3631
Washington	WA	363	746	13372	1668	2337	2944
West Virginia	WV	320	648	11352	1368	1784	2495
Wisconsin	WI	288	723	16871	2291	2952	3773
Wyoming	WY	231	442	11638	1304	1867	2449

Entries in the table are the number of observations used to calculate the job finding rate and the average wage in a state-year cell.

Table 2B
Industry-level data (SIC), cell sizes 1979-1997

		job finding rate			wage		
		min	mean	max	min	mean	max
Mining	MIN	179	642	1770	1053	1698	2896
Construction	CON	3721	6106	9114	8410	9342	10647
Lumber & wood prods, ex furniture	LUM	282	550	1058	1068	1218	1520
Furniture & fixtures	FUR	157	327	576	786	998	1231
Stone, clay, concrete, glass prods	MNR	126	321	614	766	995	1317
Primary metals	PMT	140	521	1566	985	1461	2353
Fabricated metals	FMT	223	754	1639	1693	2226	3334
Machinery, ex electrical	MAC	370	1005	2350	3237	4264	5682
Electrical machinery, equip supplies	ELC	292	890	1789	2482	3527	4735
Motor vehicles & equip	MVH	241	699	1789	1434	1952	2215
Other transportation equip	OVH	129	432	842	1316	1965	2333
Professional & photo equip, watches	PHO	88	219	397	969	1158	1397
Misc mfg industries	MMA	227	395	700	807	917	1092
Food & kindred prods	FOO	662	1173	1874	2401	3094	3960
Textile mill prods	TEX	133	393	751	779	1258	1581
Apparel & other finished textil prods	APP	447	870	1398	1199	1853	2505
Paper & allied prods	PAP	96	237	426	943	1257	1528
Printing, publishing & allied inds	PUB	372	605	830	2346	2831	3186
Chemicals & allied prods	CHE	178	381	645	1801	2251	2734
Petroleum & coal prods	OIL	15	56	106	237	336	482
Rubber & misc plastic prods	RUB	188	389	699	1123	1277	1412
Leather & leather prods	LEA	51	195	474	176	368	741
Transportation	TRA	1172	1811	2688	6460	7673	8682
Communications	COM	243	323	437	2235	2767	3327
Utilities & sanitary services	UTI	155	298	507	2135	2724	3122
Wholesale trade	WHO	943	1532	2322	5982	6757	7388
Retail trade	RET	7763	10259	13961	26903	29533	32618
Banking & other finance	FIN	421	688	927	4550	5569	6394
Insurance & real estate	INS	722	1014	1436	5276	5855	6814
Business services	BSV	1157	2290	3101	3400	5835	7560
Automobile & repair services	ASV	581	871	1281	1692	2128	2481
Personal serv ex private hhs	PSV	1031	1674	2360	3488	4274	5036
Entertainment & recreation	ENT	726	1051	1391	1658	2285	2995
Health services	HEA	1669	2274	3129	12566	15434	17922
Educational services	EDU	1243	1838	2855	14875	16391	18584
Social services	SOC	584	858	1072	2535	3389	4327
Misc professional services	MSV	644	954	1410	3697	5976	7755

Entries in the table are the number of observations used to calculate the job finding rate and the average wage in an industry-year cell. Industries are defined according to the 2-digit Standard Industrial Classification (SIC).

Table 2C
Industry-level data (NAICS), cell sizes 1998-2015

		job finding rate			wage		
		min	mean	max	min	mean	max
Mining	MIN	124	272	563	908	1206	1612
Construction	CON	2977	5179	9942	8249	9798	11839
Nonmetallic mineral product manufacturing	MNR	81	158	328	454	621	788
Primary metals and fabricated metal products	MET	322	553	1081	1790	2248	2726
Machinery manufacturing	MAC	203	396	783	1361	1856	2759
Computer and electronic product manufacturing	CEM	144	377	763	1085	1633	2145
Electrical equipment, appliance manufacturing	ELC	86	206	622	427	948	2092
Transportation equipment manufacturing	VEH	340	606	1457	2297	2715	3137
Wood products	LUM	56	196	357	249	637	960
Furniture and fixtures manufacturing	FUR	102	196	425	418	656	821
Miscellaneous and not specified manufacturing	MMA	236	439	784	1354	1525	1650
Food manufacturing, Beverage, and tobacco products	FOO	433	666	980	2376	2535	2778
Textile, apparel, and leather manufacturing	TEX	192	374	554	563	991	1962
Paper and printing	PAP	167	347	565	958	1604	2491
Petroleum and coal products manufacturing	OIL	22	40	74	186	224	265
Chemical manufacturing	CHE	136	267	579	1447	1589	1804
Plastics and rubber products	RUB	92	204	334	543	824	1221
Wholesale trade	WHO	595	977	1395	3461	5071	6565
Retail trade	RET	3653	5440	8383	16701	19049	20716
Transportation and warehousing	TRA	980	1568	2481	6516	7233	8248
Utilities	UTI	98	166	245	1555	1787	2221
Publishing industries (except internet)	PUB	80	182	389	599	853	1154
Broadcasting and Telecommunications	COM	235	476	801	1793	2296	2828
Information and data processing services	INF	52	201	907	381	1115	3040
Finance	FIN	448	821	1462	4891	5507	6104
Insurance	INS	227	394	745	2806	3102	3405
Real estate	RES	289	483	823	1955	2192	2494
Rental and leasing services	REN	28	126	298	204	434	651
Professional and technical services	PSV	1133	1881	3082	7180	8934	9919
Administrative and support services	ASV	1275	2964	5016	2705	4742	5769
Educational services	EDU	1113	2021	2952	14799	16710	17871
Hospitals	HOS	420	745	1282	7842	8930	12006
Health care services, except hospitals	HEA	510	1685	3045	5504	9676	11607
Social assistance	SOC	583	971	1561	2863	3294	3646
Arts, entertainment, and recreation	ENT	901	1254	1756	2991	3176	3688
Accommodation	ACC	501	740	1098	1875	2198	2590
Food services and drinking places	FSV	2647	3942	5614	8317	9660	10118
Other services (excl. government)	MSV	1330	1867	2735	6271	6969	7553

Entries in the table are the number of observations used to calculate the job finding rate and the average wage in an industry-year cell. Industries are defined according to the 2-digit North American Industrial Classification System (NAICS).

G Job-Finding Rates by Segment of Destination

While the construction of the aggregate job finding rate based on the Current Population Survey (CPS) is well understood (Shimer (2012), Elsby, Michaels, and Solon (2009)), it is less clear how to best measure the *segment-specific* job finding rates required for the accounting exercise in this paper.

At the core of the problem is the construction of segment-specific unemployment from CPS data. For unemployed workers the CPS only reports the last industry (or occupation) of employment, but not the segment of the labor market in which they are looking for a job.[32] In our baseline we attribute unemployed workers to the industry in which they last held a job. This is in line with the standard practice of the Bureau of Labor Statistics (BLS) and used for their publications regarding unemployment by industry (and occupation).[33] It is also the approach used in the related literature, for example, by Barnichon and Figura (2015) and Şahin, Song, Topa, and Violante (2014).[34] However, it is not consistent with most directed search models nor with the model proposed in this paper, which would attribute unemployed workers to the industry in which they are searching for a job and not to the industry in which they last held a job.

A concern is that attributing workers to their last industry of employment might impact our empirical results. For example, while the assumption that workers mostly look for work in the industry in which they used to work before losing their job might be reasonable in normal times, mobility across labor market segments in recessions is potentially higher. In particular our findings regarding the cyclicality of mismatch unemployment might therefore be affected.

In this appendix we describe an alternative method of calculating segment-specific finding rates to address these points.

G.1 Method

Using the CPS matched basic monthly data we consider the subset of unemployed workers who transitioned from unemployment to employment in an industry i in month τ. For each individual j who found a job in month τ and with month-in-sample (MISH) 2, 3, or 4 we also observe the CPS variable DURUNEMP, the number of weeks a worker has been unemployed in month $\tau - 1$. We can therefore infer the months in which individual j was not successful in transitioning from unemployment to employment in industry i.

Formally, for each individual j in industry i in month t, we define a variable uu_{ijt} equal to 1 if $t \in \{\tau - 1, \dots, \tau - k\}$, where k is unemployment duration in months, i.e. $k = \text{int}\big(1 + \frac{12}{52} \cdot \text{DURUNEMP}\big)$, and equal to 0 otherwise; and a variable ue_{ijt} equal

[32] In this note, we think of the labor market as being segmented by industries. However, the method proposed here applies when segments are operationalized as occupations.

[33] For example, Table A-14 in the Labor Force Statistics.

[34] However, Şahin, Song, Topa, and Violante (2014) provide a robustness check by applying a correction for the direction of search based on matched CPS data. The approach differs from the method presented in this note. For example, it does not make use of an unemployed worker's unemployment duration.

to 1 if $t = \tau$ and 0 otherwise. For each industry i and time t we then construct the number of individuals transitioning from unemployment to unemployment and from unemployment to employment as $\overline{uu}_{it} = \sum_j uu_{ijt}$ and $\overline{ue}_{it} = \sum_j ue_{ijt}$. Our measure of the job finding rate for industry i at time t is given by the share of successful transitions from unemployment to employment of total unemployed who eventually find a job in industry i, ${p'}_{it} = \frac{\overline{ue}_{it}}{\overline{uu}_{it}+\overline{ue}_{it}}$.

We aggregate monthly data on the job finding rates calculated in this way to annual time series by taking simple averages, and adjust the time series for the change in the reporting due to the CPS redesign in 1994. A complication is that the variable DURUNEMP is currently not available from IPUMS-CPS before the 1994 redesign. However, it can be obtained from the CPS microdata at www.nber.org/cps.

G.2 Comparison Job-Finding Rates by Destination vs Origin

Figure 6 compares the three methods used to construct segment-specific job finding rates in this paper. It provides scatter-plots of the log-finding rate $\log p_{it}$ for 37 SIC and 38 NAICS industries from 1979 to 2015 with (left-hand panel) and without (right-hand panel) deducting the industry-specific mean. In the baseline, finding rates are calculated as in Shimer (2012) ("Shimer"). The alternative method based on transitions between unemployment and employment using monthly matched CPS data is referred to as "using transitions". Both methods attribute a worker to the last industry of employment. The method presented in this appendix is referred to as "by destination." When comparing the baseline method and the method described here, the average correlation between $\log p_{it}$ by sector is 58% for the 37 industries based on the SIC classification for the 1979-1997 period and 76% for the 38 industries based on the NAICS classification for the 1998-2015 period.

The similarity between the new method and the baseline method is highest for large industries. The reason is that the larger the industry, the more likely that a worker who previously worked in that industry also looks for a new job in the same industry. Another factor is that differences due to measurement error are reduced. Figure 7 shows examples of the time-series for a very large industry (Construction), a middle-sized industry (Chemical manufacturing), and a small industry (Electrical equipment). Figure 8 shows the pattern in a more systematic way by plotting the correlation between the job-finding rates by industry of origin and destination against the size of the industry (log of employment in 2010).

Table 1 in the main text documents that the method described here also leads to similar results both in terms of the magnitude as well as regarding the cyclicality of mismatch unemployment.

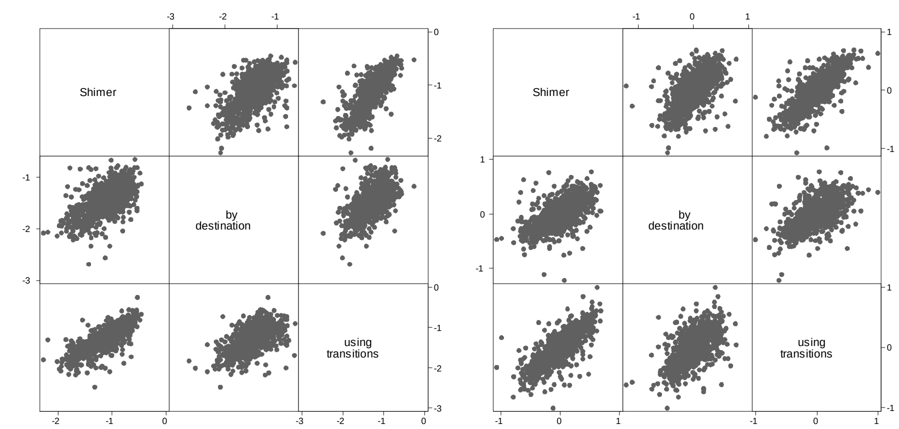
G.3 Discussion

While the new method proposed in this note does not substantially affect the findings documented in this paper, the differences might be more substantial for other empirical applications, especially since attributing a worker to the segment where he/she eventually found a job is more in line with most directed search models than relying on the industry of last employment.

The method proposed in this note also addresses another shortcoming of attributing unemployed workers to their last industry of employment. By construction the latter approach excludes inexperienced unemployed who previously never held a job as well as workers that reentered the unemployment pool after being inactive (not in the labor force, NILF) for a substantial time. The reason is that for both groups of unemployed information on the last industry (or occupation) of employment is not available in the CPS. While this particular point is of less importance for the research question in this paper, it is potentially of high relevance for other applications.

A shortcoming of the method presented here is that it implicitly assumes that a worker who a found a job at time t in industry i was looking to work in that industry during the whole unemployment spell while in reality he/she might have kept redirecting his/her search. There are scenarios where this can lead to biased results. For example, an unemployed worker might first try to find a job in his "preferred" industry and only when not successful settle for a job in a less attractive industry. That said, the proposed method is still superior in this regard compared to allocation all workers to their segment of origin.

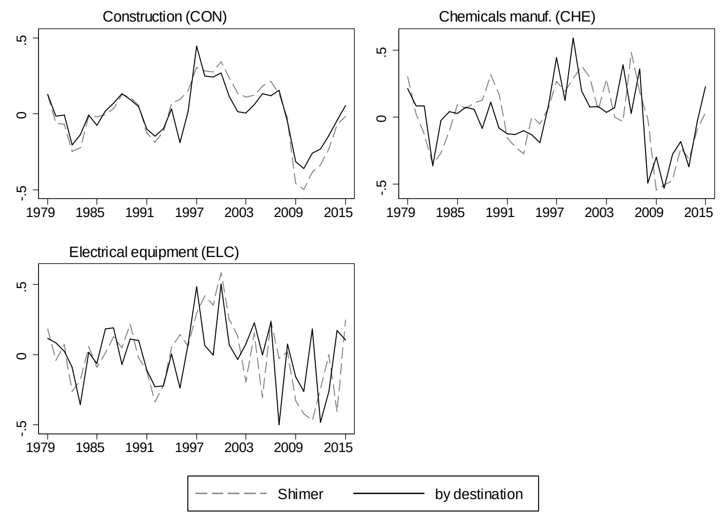
Figure 6
Scatterplots of segment-specific job finding rates



On the left scatter-plots of the log-finding rate $\log p_{it}$ for 37 SIC and 38 NAICS industries from 1979 to 2015 are shown. On the right scatter-plots are shown when the industry-specific mean is deducted: $\widehat{p}_{it} = \log p_{it} - \sum_{t=1}^{T} \log p_{it}$.

Figure 7

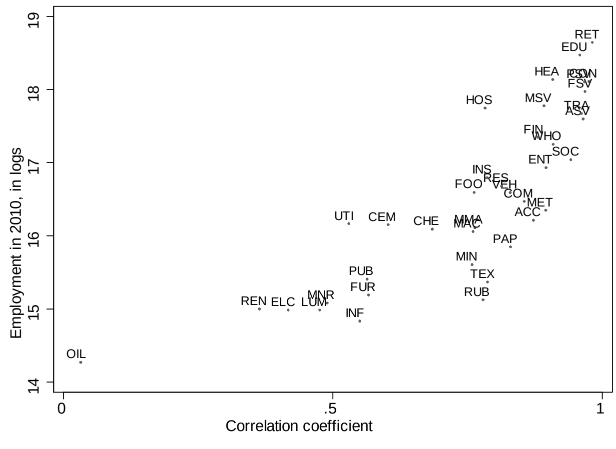
Job finding rates for Construction, Chemical manufacturing, and Electrical equipment



This compares $\widehat{p}_{it}$ using the baseline approach and the new method over time for three industries. The correlation coefficient between the two measures is 0.91, 0.68, and 0.5 for Construction, Chemical manufacturing, and Electrical equipment, respectively. Note that there is a breaks in the exact definitions of industries “CHE” and “ELC” between 1997 and 1998, see Table 2B and 2C in the main text.

Figure 8

Correlation coefficient and industry size



This shows a scatter-plot of the correlation coefficient between $\widehat{p}_{it}$ calculated according to the baseline method and the new method proposed in this note against industry size as measured by employment in year 2010 (in logs) for 38 NAICS industries.