Labor Market Selection and the Dynamics of a Recovery

Lukas Mann*

January 27, 2025

Abstract

This paper explores the role of selection in shaping the dynamics of unemployment during recoveries. A matching model with many-to-many matching and permanent worker heterogeneity delivers such selection and generates recovery unemployment dynamics that mirror the data closely. In line with empirical evidence, the model predicts that, during a recession, firms become more selective and job finding rates decline more for less productive, unemployed workers. This reinforces negative composition effects and creates a feedback loop, which slows down the recovery. I find empirical support for the cyclicality of job seeker quality implied by the model in data from the NLSY.

*Federal Reserve Bank of Minneapolis. Email: lukas.mann@mpls.frb.org. The views presented in this paper are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. I thank Gianluca Violante, John Grigsby, Richard Rogerson, Gregor Jarosch, Kieran Larkin, María Alexandra Castellanos, two anonymous referees, and various seminar participants for their constructive feedback. Edited by Greg Kaplan.

Introduction

Over the past 50 years, recessions in the US labor market have been characterized by a recurrent pattern: Unemployment rises within a short span of time and then follows a slow downward trend during the recovery. As pointed out recently by Ferraro (2023), the convergence speed and shape of these recovery paths present a puzzle to the literature studying unemployment fluctuations, as existing matching models typically fail to generate realistic unemployment dynamics during a recovery. In such models, an increase in the number of unemployed workers substantially improves matching prospects for firms, which leads to an instantaneous jump in the number of posted vacancies. This vacancy inflow pushes transition rates to their steady state levels, and any excess unemployment dissipates within months.

In this paper, I develop an equilibrium model of labor market selection that is able to generate realistic unemployment dynamics during a recovery. In the model, the relatively slow adjustment of unemployment is the outcome of a feedback mechanism. Two assumptions are key: First, workers are ex-ante heterogeneous. Second, they compete with each other for the same jobs and are selected according to a common ranking. This selection helps to generate the correct recovery dynamics through the following mechanism: Early during a recovery, labor markets are slack. Under slack markets, more applicants compete for the same vacancy, and low-productivity workers are less likely to match successfully. This shifts hiring towards higher-quality workers who are more likely to be employed but skews the pool of job seekers towards lower-quality workers. The qualitative decline of the pool of job seekers in turn reduces the incentive to post vacancies, which keeps markets slack. As a result, the recovery drags on, and the unemployment rate converges to steady state very slowly, in line with aggregate data.

I present evidence from the National Longitudinal Survey of Youth 1979 (NLSY) that establishes the empirical plausibility of this mechanism. As is consistent with the model, workers' ranks in the worker quality distribution can be identified from their lifetime job finding probabilities. I show that workers of low rank have job finding probabilities that are markedly more procyclical than those of high rank workers. I also document procyclical composition effects in the NLSY. In particular, I show that taking into account search of workers outside of unemployment, as well as accounting for a worker's likelihood of selection, markedly increases the procyclicality of the quality of job seekers relative to a measure that focuses on the average quality of unemployed workers.

In order to model the selection process, this paper introduces a model of worker-firm matching that allows for multiple encounters on both the worker and the firm side and features permanent worker heterogeneity. The combination of these two features gives rise to worker selection on the firm side; that is, workers compete for the same jobs but differ in quality and therefore in their likelihood of getting selected by firms. In this framework, cross-sectional differences in UE cyclicality arise naturally: A high rank worker typically needs only one single encounter to get hired, whereas a low rank worker needs to be close to the only viable candidate for a position. During times of slack markets, when vacancies receive more encounters, the chances of getting hired thus decline more for low rank workers than for high rank ones. As a result, the pool of job seekers shifts towards lower quality applicants. This reduces the incentive to post vacancies, which creates a feedback loop. At the same time, slack markets also allow the firm to be more selective about its applicant pool and raise firm values through increased selectivity rather than through increased matching rates, which further slows down the recovery relative to that of a model with a standard one-to-one matching function.

When taken to the data, the model captures the empirically observed slow speed of recent recoveries. Moreover, conditional on an exogenous path of transition rates into and between non-participation and unemployment, the shape of the recovery path of unemployment is almost perfectly replicated by the model for each of the six most recent recessions in the data. I find that the model succeeds in doing so by capturing the empirical co-movement of transition rates between employment and unemployment - specifically, a sluggish recovery of the UE rate over time.

The work presented here relates to several different strands of literature. The issue of sluggish post-recession unemployment adjustment has been studied in several papers, such as Hall and Kudlyak (2022), Ferraro (2023) and Gregory, Menzio and Wiczer (2024) (GMW). Like me, GMW employ a model with heterogeneous workers to study unemployment dynamics during the Great Recession. GMW consider workers that differ in their average job finding and separation rates and study composition effects as a force that can slow down the adjustment of unemployment under persistent productivity shocks. In contrast, I highlight a fundamentally different selection-based mechanism, which causes a general equilibrium feedback loop that operates through composition effects and thus generates slow recoveries even if productivity immediately returns to steady state. Hall and Kudlyak (2022) discuss a wide variety of mechanisms that have the potential to explain the speed and shape of recoveries in the data. Among them, composition-based effects stand out as one of the leading approaches to achieve model outcomes that match the data. This idea is somewhat controversial in the literature, and there is an active debate on the cyclicality of the composition of the unemployed (see Barnichon and Figura (2015) and Mueller (2017) for two papers on opposite sides of this issue). In my model, recoveries slow down primarily through composition effects, but strong composition effects can occur with little cyclicality in the average quality of unemployed workers. Using data from the NLSY, I provide new empirical evidence that the cyclicality of job seeker quality implied by the model is supported empirically.

One of the key challenges in creating realistic recovery outcomes is to make the unemployment rate sufficiently detached from its long-term steady state. Therefore, this paper also relates more broadly to the large literature studying amplifiers of unemployment fluctuations, initiated by Shimer (2005) and Hagedorn and Manovskii (2008) and more recently revisited by Hall (2017), Eeckhout and Lindenlaub (2019), and Mercan, Schoefer and Sedláček (2021), among others. Pries (2008) investigates the potential of worker heterogeneity to generate persistent unemployment fluctuations and finds that even with worker heterogeneity, much unemployment persistence remains to be explained. A contribution of this paper is to show that conditional on separations, worker heterogeneity can fully account for persistent unemployment dynamics, once selection is also taken into account. In addition, my model not only matches the persistence of unemployment but hits the shape of the unemployment series over the course of each recovery very well. The paper closest to this idea is perhaps Ferraro (2018), which studies the skewness of the long-run unemployment distribution in addition to its variance, although the focus is on matching long-run aggregate statistics rather than individual recoveries. The model in Ferraro (2018) also features a form of selection: Some workers become unemployable in some aggregate states. In contrast, the way in which selection is modeled in this paper offers an arguably more realistic description and micro-foundation of individual job finding rates as a function of worker quality and the aggregate state, and allows me to match empirical cross-sectional differences in the level and cyclicality of job finding rates.

The ranking model introduced here is similar in spirit to the one developed by Blanchard and Diamond (1994), who study a model in which vacancies meet with several workers and the one with the lowest recorded unemployment duration is chosen. Their paper, however, emphasizes the role of ranking for wages, not for the dynamics of unemployment. In fact, the ranking assumption in their model has no effect on the behavior of aggregate unemployment and carries implications only for wages and individual unemployment duration. In contrast, ranking as conceptualized in this paper has major allocative consequences: During slack markets, the pool of job seekers worsens in quality, and hiring shifts away from the unemployed and towards the employed.

In allowing for multiple encounters per vacancy, my matching stage has parallels with other models with multiple encounters per worker or firm, such as the ones in Barnichon and Zylberberg (2019) and Bradley (2022) which are adaptations of the urn ball model in the classic paper by Butters (1977). In contrast to these papers, the matching stage in my model features multiple encounters on both the worker and the firm side. Both Wolthoff (2018) and Birinci, See and Wee (2024) consider models that feature two-sided multiplicity of encounters but do not feature permanent worker or firm heterogeneity. In contrast, my model features permanent heterogeneity on the worker side, which drives the selection mechanism. Although not analyzed here, the matching setup does in principle allow for permanent heterogeneity on the firm side as well, and I discuss possible applications of this idea.

The idea that the unemployed have on average more difficulty finding a job than the employed also relates to Engbom (2021), which studies a model in which the expected value of a match falls with a rise in the number of unemployed workers, since they apply for less suitable jobs. The model presented here highlights that the job finding rates of unemployed workers can be comparatively low for a different reason: The unemployed simply tend to be less productive workers and consequently will be the top candidate for a position less frequently. In matching this feature of the data, the model is therefore able to generate endogenously the finding in Faberman et al. (2017) that the employed receive more matches per unit of search effort than the unemployed. In addition, the model predicts that the wage premium of the typical hire from employment relative to a hire from unemployment is 9 log points, which accounts for a quarter of the total difference identified in Faberman et al. (2017).

The rest of the paper proceeds as follows: Section 1 reviews the dynamics in transition rates that are required to generate realistic recoveries and discusses why the textbook model fails to generate these dynamics. Section 2 presents evidence on compositional effects and cross-sectional differences in the level and cycliality of transition rates in the NLSY. Section 3 outlines the model, and Section 4 discusses its calibration. Section 5 uses the model to simulate recoveries and investigates the forces driving the observed dynamics. Section 6 concludes.

1 Recovery dynamics in the data

In US data, the unemployment rate exhibits a reliable pattern of a slow downward glide back to steady state after each recession, illustrated by Figure 1. With the exception of the

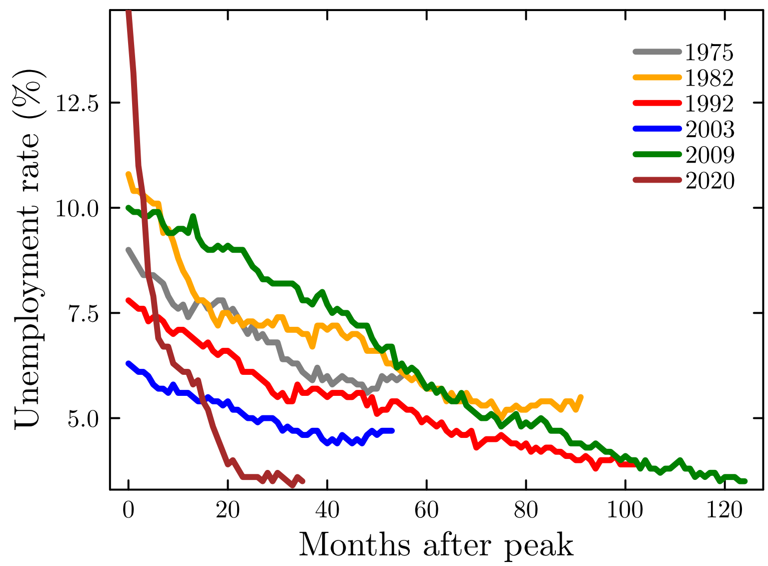


Figure 1: Unemployment paths in the last 6 recovery episodes

Notes: Unemployment rates during recovery episodes according to the Bureau of Labor Statistics (BLS). Covered episodes: May 1975 - Dec 1979, Dec 1982 - Jul 1990, Jun 1992 - Dec 2000, Jun 2003 - Nov 2007, Oct 2009 - Feb 2020, Apr 2020 - Mar 2023.

post-COVID recovery, each recovery path of unemployment is quite similar in shape and speed. This feature, however, has become a puzzle in the literature because the textbook search model fails to produce such recovery dynamics.

To see why the textbook model fails, consider the following equation describing continuous-time unemployment dynamics in a broad class of labor market models with only two states, employment and unemployment:

$$\dot{u}_t = -\lambda_t u_t + \delta_t (1 - u_t). \tag{1}$$

Here, λ_t is the average rate at which the unemployed find a job, δ_t is the separation rate into unemployment, and u_t is the unemployment rate. Consider now a scenario in which these transition rates are fixed over some time period (i.e., $\lambda_t = \lambda$ and $\delta_t = \delta$). In this case, equation (1) has the solution

$$u_t = e^{-(\lambda+\delta)t}\left(u_0 - \frac{\delta}{\lambda+\delta}\right) + \frac{\delta}{\lambda+\delta}. \tag{2}$$

The half-life of excess unemployment arising from this equation, a convenient expression for the speed at which excess unemployment dissolves, is $\frac{\log 2}{\lambda+\delta}$. Using monthly values of $\delta = 0.015$ and $\lambda = 0.25$, which are motivated by typical ranges found in CPS data, this half-life is 2.6 months. This high "natural" rate of convergence of the unemployment rate stems from the high turnover of unemployment implied by average UE and EU rates and means that the long-run unemployment rate implied by current transition rates is typically a good approximation of the true current unemployment rate u_t; that is,

$$u_t \approx u_t^* = \frac{\delta_t}{\lambda_t + \delta_t} = \frac{1}{1 + \frac{\lambda_t}{\delta_t}}.$$

Importantly, the law of motion described by equation (1) is an identity that holds very generally across many environments, including models with worker heterogeneity. For this reason, $u_t \approx u_t^*$ holds in any two state model with realistic average UE and EU rates. Any good model of recoveries must therefore produce realistic movements in u_t^* itself. This implies that transition rates need to move in a way that is close to their empirically observed patterns. Concretely, the UE/EU ratio $x_t = \lambda_t/\delta_t$, which is a sufficient statistic for $u_t^* = \frac{1}{1+x_t}$, must follow the correct dynamics.

It turns out that the textbook Diamond-Mortensen-Pissarides (DMP) model fails this test. The DMP model implies an immediate reversal of x_t to steady state values once a shock has disappeared, which generates an almost immediate recovery. In this model, a free entry condition yields λ as a direct function of fundamentals (e.g., productivity). Thus, λ returns to steady state when fundamentals revert to steady state. Under the assumption that $\delta_t = \delta$ is fixed and exogenous, this implies $u_t^* = u_{ss}$ throughout the recovery. Intuitively, during times of high unemployment, vacancies are likely to encounter many applicants, which increases vacancy posting up to the point where the vacancy filling rate returns to steady state. This happens to also be the point at which the job finding rate is at steady state and therefore $u_t^* = u_{ss}$. Unemployment evaporates within a short time, in line with equation (2).

What are the empirical patterns of x_t that a successful model of recoveries would have to capture? Figure 2 provides the answer. It shows the empirically observed values of

$\log x_t$ over the course of each recovery since 1975. The orange line traces out the realized path of the log UE/EU ratio, whereas the green and red lines decompose $\log x_t$ into its constituent UE and EU components, using the fact that

$$\log x_t = \log UE_t - \log EU_t$$
$$\implies \log x_t - \log x_{ss} = (\log UE_t - \log UE_{ss}) + [-(\log EU_t - \log EU_{ss})], \tag{3}$$

which yields an additive decomposition of the dynamics of $\log x_t$.

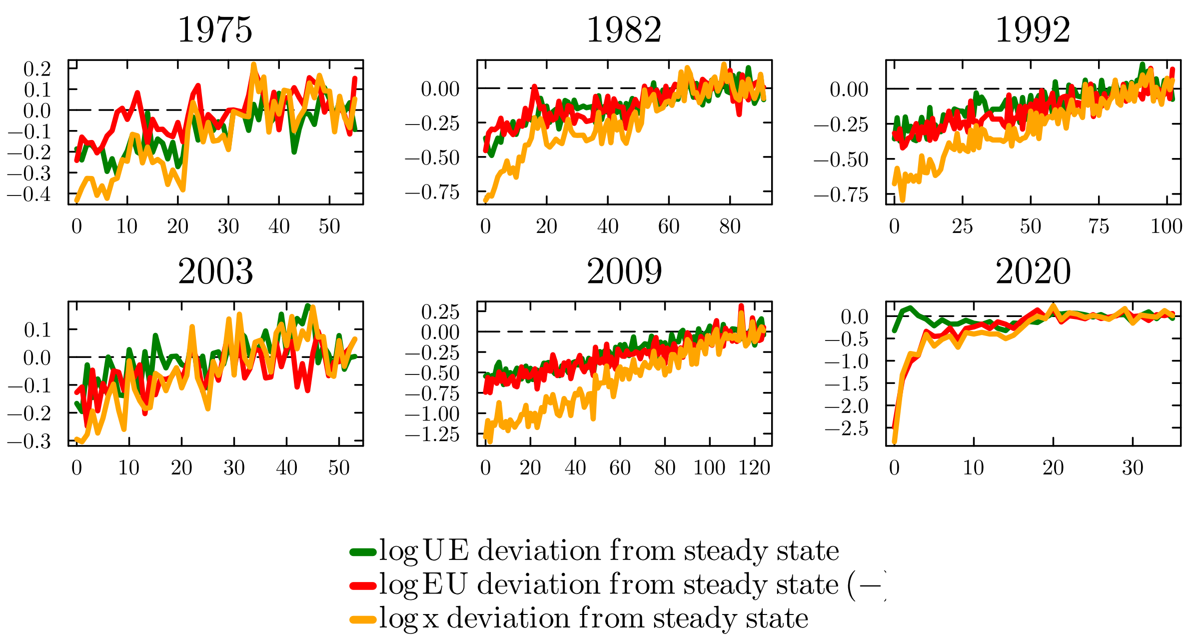


Figure 2: $\log x_t$ and its additive decomposition during the last 6 recoveries

Notes: Decomposition of the log UE/EU ratio according to equation (3). UE and EU rates are the gross flow rates of all workers in the CPS, computed from the basic CPS files as in Shimer (2012). Covered episodes as in Figure 1. Steady state values are computed as the mean of the final 6 months in each episode. The green line traces out $(\log UE_t - \log UE_{ss})$, the red line traces out $[-(\log EU_t - \log EU_{ss})]$, and the orange line traces out $\log x_t - \log x_{ss}$.

With the exception of the post-COVID recovery, which is driven mostly by movements in separation rates, the dynamics of x_t appear very similar across episodes: About half of the deviation of $\log x_t$ from its steady state is driven by movements in $\log UE_t$ and about half by movements in $\log EU_t$. The key to generating empirically realistic recovery dynamics is thus to get the same slow and steady increase of $\log UE_t \approx -\log EU_t$ over the course of the recovery.

To give a completely satisfying explanation for the underlying source of recovery dynamics, one would have to consider a model that produces both endogenous separations *and* job finding rates that co-move with separations as shown in Figure 2. The aims of this paper will be less ambitious than this challenging benchmark. Instead, we will show that a model with selection successfully generates the correct movement in the UE rate *given* the EU rate. In this sense, the model can account for about half of the underlying puzzle. The question of how to generate realistic movements in $\log EU_t$ is left for future research.

The key idea of the model presented below is that selection effects translate generally slack markets into environments with a job finding rate that is depressed, particularly for low-rank workers. These workers are more likely to search from unemployment, which depresses the unemployment exit rate. Their continued search depresses the quality of job seekers, which discourages vacancy posting. This, in turn, keeps markets slack, generating a feedback loop. The job finding rate of the unemployed thus recovers only gradually and reproduces the co-movement between the EU and UE series shown in Figure 2. Thus, the model produces recoveries that match both the speed and the shape of recoveries in the data.

The mechanism through which the model can achieve this crucially hinges on two ideas. First, composition effects are a major force that disincentivizes hiring during the early stages of a recovery. Second, less productive workers are comparatively more affected by slack markets and thus fuel these composition effects. The next section documents empirical support for both these ideas.

2 Empirical evidence

To motivate the model and provide some empirical evidence on the central mechanism of this paper, I use data from the National Longitudinal Survey of Youth 1979 (NLSY). The NLSY is an ongoing survey that has been following the same cohort of workers over their working lives since 1979. I use the representative sample of workers, dropping the military and supplemental samples, and consider only workers of at least 25 years of age. This leaves 6111 workers in the sample, with work histories recorded between 1982 and 2019. Workers are interviewed yearly or bi-yearly and report their working status and wage for every week since the previous interview. I conduct the analysis on a monthly level, classifying workers as "employed" when they are recorded as working in any given week during a month, "unemployed" if they are recorded as unemployed at any point during a given month (but not as employed), and "non-participating" otherwise.

2.1 External validity of unemployment dynamics in the NLSY

I begin by checking whether unemployment dynamics in the NLSY are representative for the economy at large. That is, I test whether unemployment rates in the NLSY are comparable with those reported by the Bureau of Labor Statistics (BLS). Figure 3 compares both series. Unemployment rates of the NLSY cohort are highly correlated

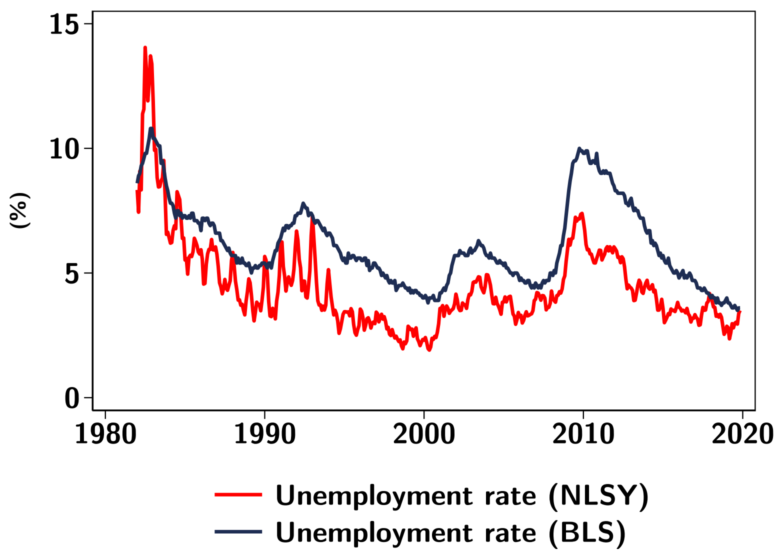


Figure 3: The unemployment rate according to BLS and NLSY

Notes: Dark blue line: Monthly unemployment rate according to the BLS. Red line: Monthly unemployment rate in the NLSY, defined as the ratio of the number of unemployed workers over the sum of employed and unemployed workers in a given month.

with aggregate unemployment. Unemployment for young workers in the beginning of the sample is generally higher than at the end of the sample, when workers are older. The unemployment rate of the NLSY cohort trends down over time, stabilizes at a level below the BLS unemployment rate in the mid-1990s, and tracks the BLS cyclicality throughout the sample.

2.2 Sensitivity of UE and EU rates to the cycle

We begin the analysis by studying the differential exposure of workers of different types to the cycle. The NLSY tracks workers throughout their entire working life, which means that it is possible to compute a given worker's lifetime job finding rate

$$\mathrm{LJR}_i = \frac{\sum_t \mathbb{I}(\mathrm{Unemployed})_{i,t}\mathbb{I}(\mathrm{Employed})_{i,t+1}}{\sum_t \mathbb{I}(\mathrm{Unemployed})_{i,t}}$$

which is defined as the share of months the worker spends in unemployment that are followed by a month in employment. A worker's rank in the aggregate distribution of LJRs offers a natural notion of "quality" or "rank" of that worker; that is, it captures how desirable that worker is, from the perspective of a firm, compared to other potential applicants. Consequently, we will henceforth refer to a worker's quantile in the aggregate distribution of LJRs as their "rank" or "quality." This notion of rank is consistent with the model outlined in Section 3.

To compare worker types across the quality distribution, I bin all workers in the NLSY into one of five quintiles q_5 based on their LJRs. I then calculate the average monthly job finding rate in every year for each group q_5. Figure 4 shows log job finding rates by q_5.

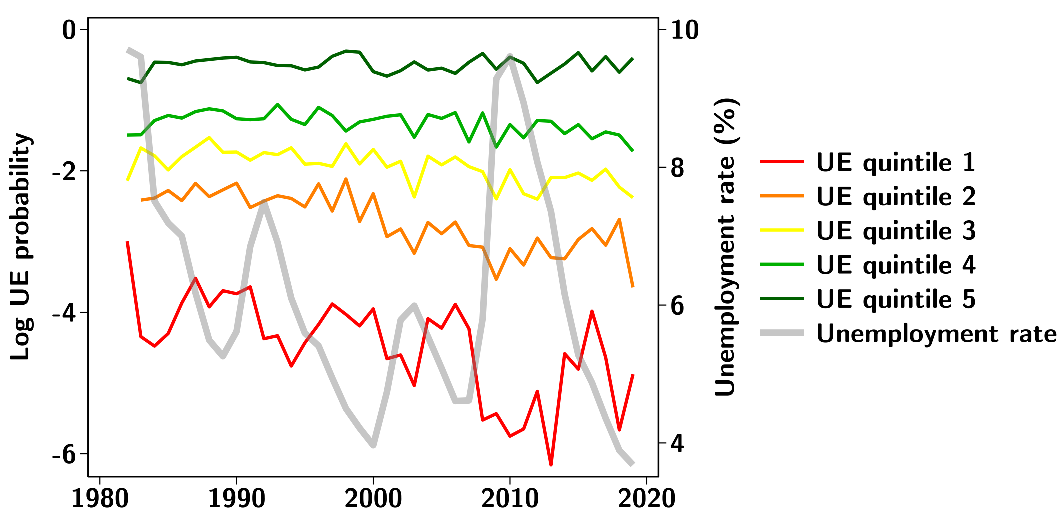


Figure 4: UE probability over time by worker quintile

Notes: Log of the average yearly UE probability by LJR quintile q_5 in the NLSY (left scale). Shown in gray is the BLS unemployment rate (right scale) as a measure of the state of the business cycle.

Job finding rates are visibly more cyclical at the bottom ($q_5 = 1$) than at the top ($q_5 = 5$). Denoting the year as t and the average yearly BLS unemployment rate as UR_t, we run the following regression to confirm this finding:

$$\log \mathrm{UE}_t^{q_5} = \beta_0^{q_5} + \beta_1^{q_5} \log \mathrm{UR}_t + \gamma_1^{q_5} t + \gamma_2^{q_5} t^2 + \varepsilon_t^{q_5}. \tag{4}$$

The first row of Table 1 collects the coefficient $\beta_1^{q_5}$ for each quintile q_5. The regression confirms that lower rank workers' job finding rates are more exposed to the business cycle than those of high rank workers. The job finding rates of low rank workers drop more during times when unemployment is high. This fact is relevant for the aggregate dynamics of unemployment in at least two distinct ways.

First, because a decline in the aggregate job finding rate is more pronounced for low-rank workers and because these workers are disproportionately unemployed, a decline in aggregate labor market tightness should lead to a disproportionately large decrease in the rate at which the unemployed find jobs. In comparison, employed workers' job finding rates should not be as cyclical, as the pool of employed workers contains more workers with a high rank. Section 5 provides evidence that this is indeed observable in aggregate data from the CPS.

Second, the composition of the pool of job seekers is driven by the relative cyclicality of EU and UE rates by each group. If differences in the cyclicality of EU rates are sufficiently small across groups, then the differences in UE cyclicality can lead to composition effects that may discourage hiring.

To check how much EU rates vary across groups, we repeat the earlier regression, this time for EU rates by q_5:

$$\log \mathrm{EU}_t^{q_5} = \delta_0^{q_5} + \delta_1^{q_5} \log \mathrm{UR}_t + \eta_1^{q_5} t + \eta_2^{q_5} t^2 + \nu_t^{q_5}. \tag{5}$$

The second row of Table 1 collects the results. Differences in EU cyclicality are small across groups, but interestingly, lower rank workers are somewhat more exposed to the cycle than high rank workers are. This finding stands in contrast to the findings of Mueller (2017), which provides evidence that separations in the CPS are more cyclical for high-wage workers than for low-wage workers. The notable difference between Mueller's approach and the approach pursued in this paper lies in the classification of workers: I classify workers not by their wage but by their lifetime job finding rate. This approach is consistent with the idea of selection considered in this paper: When two workers compete for the same job, the more productive candidate should observe a higher job finding rate. However, if there are multiple, segregated labor markets, this conclusion breaks down: Consider two workers, A and B, that operate in distinct, segregated markets. Worker A may be more productive and have a higher wage than B but compare unfavorably with their market competitors (and thus have a low LJR), whereas worker B may be the most qualified candidate in their respective market (and thus have a high LJR). The exercise conducted here would assign a higher quality to worker B. In contrast, Mueller (2017) would classify worker A as a more desirable hire from a firm perspective.

2.3 Composition of the pool of job seekers

The large observed differences in UE cyclicality across worker groups suggest that the composition of the pool of job seekers may shift towards lower quality workers in recessions. We now evaluate this hypothesis directly in the NLSY. To do this, a more granular notion of worker quality is useful. To that end, I group workers into centiles q_{100} according to their lifetime job finding probability. We will refer to these centiles as the worker's "rank" or "quality" in what follows. I then calculate the average worker centile among the employed, unemployed, and non-participating population, denoted respectively by $\bar{q}^e_{100}$, $\bar{q}^u_{100}$, and $\bar{q}^n_{100}$. Figure 5 plots the resulting measure of average quality by employment status. The quality of the pool of employed workers is steady over time and shows little cyclicality. The same is true for the pool of non-participating workers. The average quality of the pool of unemployed is characterized by a general downward trend over the sample period but also exhibits fluctuations at business cycle frequencies.

We begin by focusing on the pool of unemployed workers. To check whether fluctuations in the average quality of the unemployed are pro- or countercyclical, Figure 6 shows the average rank of unemployed workers along with the aggregate unemployment rate.

Two different patterns emerge before and after 1995: In the period before 1995, the quality of the unemployed appears countercyclical. In the period after 1995, and particularly around the Great Recession, the pattern is procyclical.

The average quality of unemployed workers, or variants thereof, are commonly used to study composition effects. However, if one is interested in measuring shifts in worker quality composition that affect firm hiring incentives, there are at least two shortcomings of this measure: First, not all job seekers are unemployed; a large share of hires comes from the employed and non-participating population. Therefore, the worker quality composition that is relevant from a firm perspective is a weighted average of the quality of

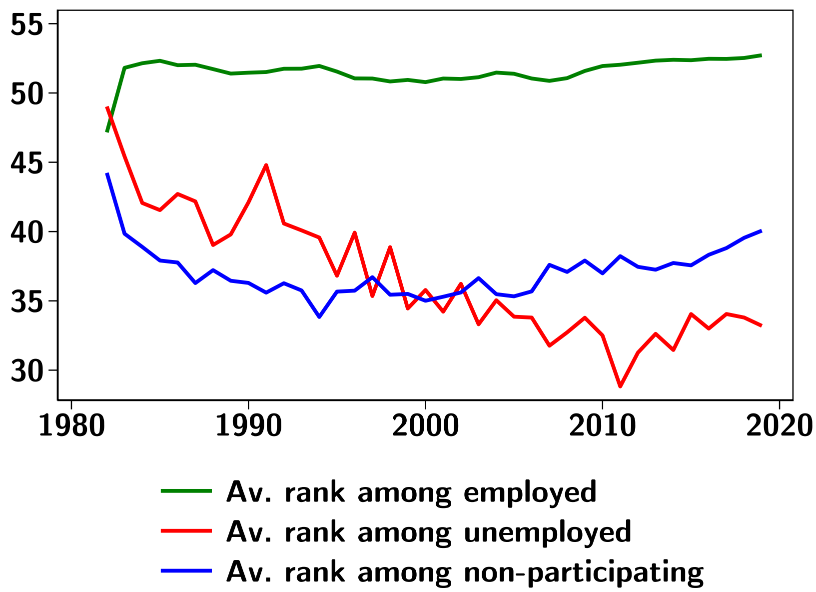


Figure 5: Average rank of workers by employment status

Notes: The green, red and blue lines track the yearly average LJR centile of workers in the pool of employed ($\bar{q}^e_{100,T}$), unemployed ($\bar{q}^u_{100,T}$), and non-participating workers ($\bar{q}^n_{100,T}$), respectively, for each year T.

employed, unemployed, and non-participating workers. Second, not all encounters are equally relevant to the firm, since firms should care more about the quality of workers that they are likely to hire.

In what follows, we will account for both of these ideas. We begin by measuring the composition of the pool of all job seekers, rather than that of just the unemployed, which is constructed as follows:

$$\text{SearchProb}_{i,t} = \begin{cases} 1 & \text{if } i \text{ is unemployed at time } t, \\ s_e & \text{if } i \text{ is employed at time } t, \\ s_n & \text{if } i \text{ is non-participating at time } t, \end{cases}$$

$$\bar{q}_{\text{job seekers},T} = \frac{\sum_{i,t\in T} q_{100,i}\text{SearchProb}_{i,t}}{\sum_{i,t\in T}\text{SearchProb}_{i,t}}, \tag{6}$$

where i is the worker index, $t \in T$ are different months within the same year T, and s_e and s_n are the relative search intensities of employed and non-participating workers. We use search intensities close to our model estimates ($s_e = 0.03, s_n = 0.12$). Figure

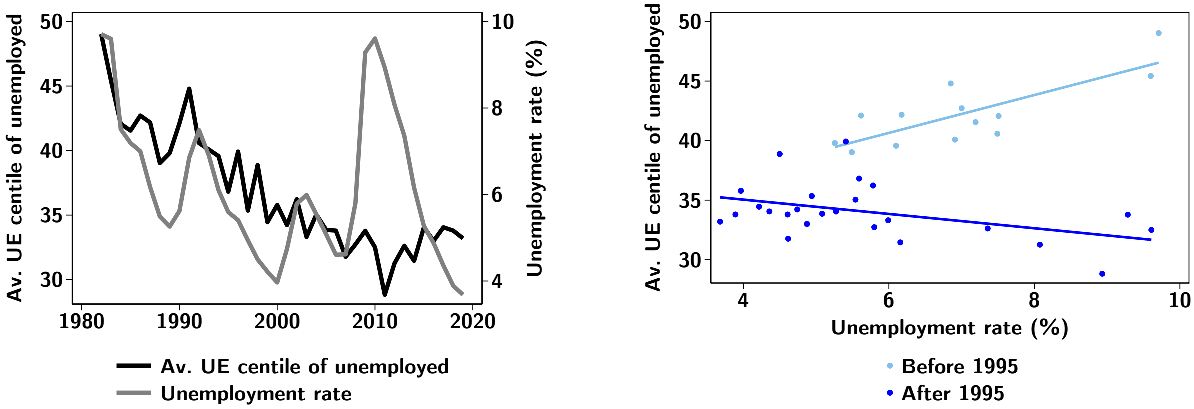


Figure 6: Average rank of the unemployed over the cycle

Notes: Left panel: The black line shows the average LJR centile of workers in unemployment ($\bar{q}^u_{100,T}$) by year T. The gray line shows the average BLS unemployment rate by year. Right panel: The relationship between the two variables is displayed as a scatter plot with color coding indicating whether a pair of observations was recorded (strictly) before or (weakly) after 1995; lines of best fit are shown for each of these two episodes (1982-1994, 1995-2018).

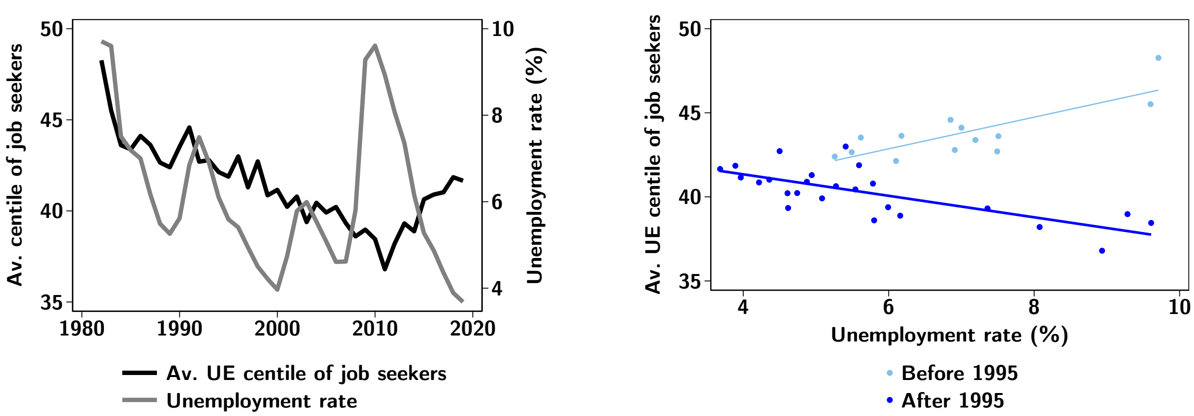


Figure 7: Average rank of all job seekers over the cycle

Notes: Left panel: The black line shows the average LJR centile of job seekers ($\bar{q}_{\text{job seekers},T}$) by year T, as defined in equation (6). The gray line shows the average BLS unemployment rate by year. Right panel: See Figure 6.

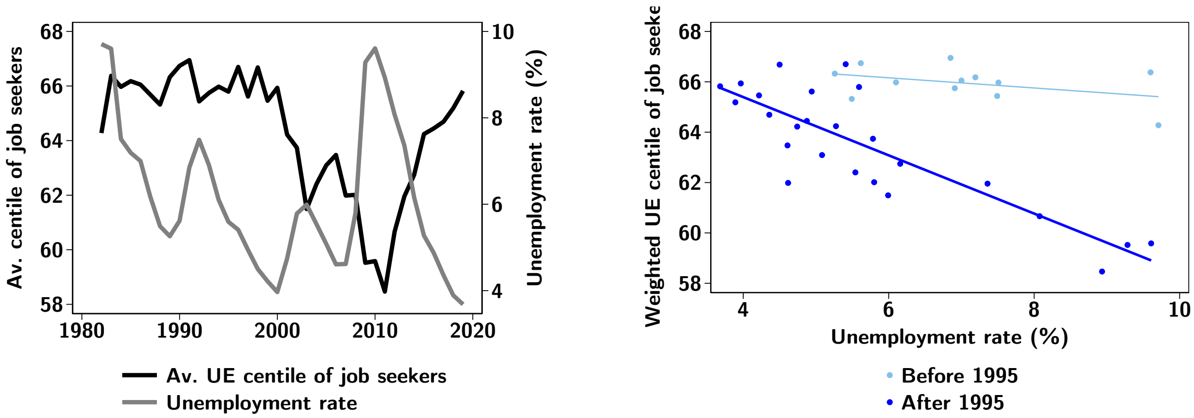


Figure 8: Weighted rank of all job seekers over the cycle

Notes: Left panel: The black line shows the weighted average LJR centile of job seekers ($\bar{q}_{\text{weighted},T}$) by year T, as defined in equation (7). The gray line shows the average BLS unemployment rate by year. Right panel: See Figure 6.

7 documents that this measure of composition is markedly more procyclical than the quality of the unemployed, partly because the worker composition shifts towards the pool of unemployed in a recession and the unemployed have lower quality on average than the employed.

It is important to note that these estimates of search intensity, which come from the model, are very conservative compared with other estimates of search intensity in the literature. This is due to the selection mechanism present in the model: Employed workers are predicted to be selected more often, and thus low search intensities are needed to justify empirically plausible EE rates. If s_e is increased to reflect estimates close to the traditionally estimated range (in models without the selection mechanism present here), the quality of the pool of job seekers becomes more procyclical. A similar effect can be observed when, as is the case below, one accounts for the differential likelihood of selection across different worker types.

As a final exercise, we account for differential likelihood of selection. In an environment with selection, not all encounters are equally important to the firm. Rather, the firm gives a larger weight to the quality of workers it is more likely to select. A worker that is twice as likely to be selected will be given a weight twice as high. As will be the case in the model, we can use the relative UE rates of different worker types as a measure of relative likelihood of selection. We thus project the 2019 log UE rate by q_{100} onto a second order polynomial in q_{100} and use this regression to predict the relative likelihood of selection:

$$\widehat{\log \mathrm{UE}}^{2019}_{q_{100}} = \xi_0 + \xi_1 \cdot q_{100} + \xi_1 \cdot {q_{100}}^2,$$

$$\mathrm{Weight}_{i,t} = \exp(\widehat{\log \mathrm{UE}}^{2019}_{q_{100,i}}) \cdot \begin{cases} 1 & \text{if } i \text{ is unemployed at time } t, \\ s_e & \text{if } i \text{ is employed at time } t, \\ s_n & \text{if } i \text{ is non-participating at time } t, \end{cases}$$

$$\bar{q}_{\text{weighted},T} = \frac{\sum_{i,t \in T} q_{100,i} \mathrm{Weight}_{i,t}}{\sum_{i,t \in T} \mathrm{Weight}_{i,t}}. \tag{7}$$

Note that we are holding the likelihood of selection fixed at the same level over time. We do so in order to distinguish changes in the composition of job seekers from changes in the likelihood of selection.

Figure 8 shows that incorporating differential selection probabilities further increases the procyclicality of our measure of average worker quality. The series of worker quality in the population of all job seekers is now markedly procyclical in the post-1995 period and even somewhat procyclical in the pre-1995 period. The quality of workers among the pool of job seekers is thus high in periods of low unemployment and low in periods of high unemployment.

Taken together, this evidence establishes that composition effects plausibly play a role in discouraging hiring during a recession and may slow down the recovery process. Cross-sectional differences in the cyclicality of job finding rates can drive such composition effects but are absent from standard models of labor markets.

In what follows, we will develop a labor search model that features selection; that is, it takes seriously the idea that multiple workers often have to compete for the same job. As will be shown below, this feature endogenously generates empirically plausible cross-sectional patterns of job finding rates over the cycle. The model has predictions for the composition effects resulting from this selection mechanism and for the dynamics of labor market recoveries that can be compared to the data.

3 Model

We now outline a model of labor market selection. Motivated by the empirical evidence above, the model's matching mechanism will have the feature that workers have endogenously different job finding rates conditional on their quality, and during a recession, job finding rates endogenously decline more for workers of low rank. The model offers a micro-foundation of this pattern: When multiple workers compete for the same job, low productivity workers are more likely to be outranked by high productivity competitors. Crucially, this effect becomes much stronger under slack markets, since slack market environments involve more competition of workers for jobs.

3.1 Environment

Workers: The economy is populated by heterogeneous workers indexed by their type $i \in [0,1]$. Without loss of generality, I assume that workers are distributed uniformly across i. At the end of any period, a worker is in one of three states, non-participating, unemployed or employed. Every worker type is characterized by a tuple $(y(i), d^u(i), d^n(i))$, where $y(i)$ is their productivity, and $d^u(i), d^n(i)$ are their relative transition rates into unemployment and non-participation, respectively. We will assume that workers are selected by firms in order of their type i; that is, the firm always selects the highest type that is willing to match with them. To make this assumption consistent with firm optimality, the value that the firm can extract from a match with a type i worker must be weakly increasing in i. We will focus on a calibration of the model for which this relationship holds at any time t.

Timing: Time is discrete and infinite. Each period consists of three stages. First, some employed and unemployed workers receive a non-participation shock. If a worker receives non-participation shock, they are out of the labor force at the beginning of the next period. The probability of this shock for a given worker depends on the worker's type as specified below. Second, among workers who do not receive this shock, some receive an unemployment shock. If a worker is hit by this shock, they will be unemployed at the beginning of the next period. In the third stage, those workers who are not hit by any of these shocks participate in a matching mechanism. I refer to the third stage as the hiring or matching stage. After it has passed, hires resolve, all transitions conclude, and the number of employed, non-participating and unemployed workers is measured. Figure 9 illustrates this timing assumption.

I assume that transitions out of employment are exogenous and type-specific. During the first stage, each employed worker of type i transitions into non-participation with probability $\delta_t^{en}(i) = d^n(i)\delta_t^{en}$. Each unemployed worker transitions into non-participation with probability δ_t^{un}. During the second stage, non-participating and employed workers transition into unemployment with probabilities δ_t^{nu} and $\delta_t^{eu}(i) = d^u(i)\delta_t^{eu}/(1 - d^n(i)\delta_t^{en})$,

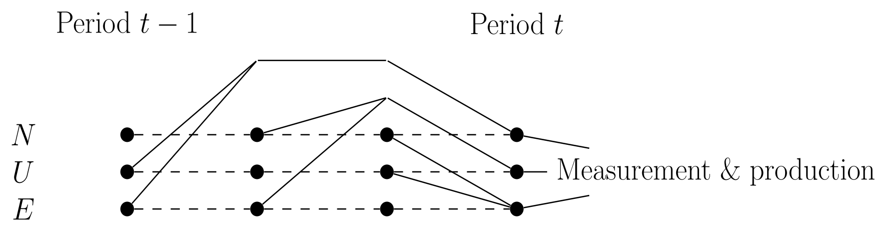


Figure 9: Timing assumptions

Notes: Timing assumptions of the model. Each column of 3 black dots stands for the N/U/E stock of workers within a given sub-period (from left to right). Black solid lines between sub-periods symbolize flows.

respectively (this makes the unconditional probability of a flow into unemployment for an employed worker at the beginning of the period equal to $d^u(i)\delta_t^{eu}$). Finally, during the matching stage, conditional on the employment state, workers become job seekers with probability 1 (if unemployed before the matching stage), s_e (if employed) or s_n (if non-participating). As a searcher, they partake in the matching mechanism described in Section 3.2. If matched, they transition into employment; otherwise, they retain their employment status.

Firms: There is a measure of atomistic and homogeneous firms. In each period, during the hiring stage, firms can decide to post an arbitrary number of vacancies at cost κ. The vacancy is then subject to the matching process described in Section 3.2 that determines whether it is filled and, if so, with what kind of worker. If successfully filled, production begins immediately after the hiring stage and continues for every period after the hiring stage until the worker stochastically separates into unemployment or non-participation.

Bargaining: Once matched, wages are set according to Nash bargaining. I assume that the outside option for any worker is always the same, irrespective of the matched job seeker's previous employment status. I assume that for any worker, their outside option is to follow the transition probabilities into non-participation or unemployment conditional on not remaining in employment. Importantly, I assume this for all matches, including those from prior employment. This assumption sidesteps the problem of non-convex bargaining sets for models with Nash bargaining and on-the-job search raised in Shimer (2006).

3.2 Matching and selection

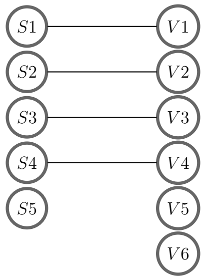
I now describe the matching stage, which is a model of the selection mechanism by which firms hire workers. It features two-sided multiplicity of encounters - a job seeker can encounter multiple vacancies and a vacancy multiple job seekers. The idea of matching games that allow for multiple encounters on the seller or buyer side goes back at least to Butters (1977) and the matching setup described below has some parallels to Butters's urn ball model. Appendix B.1 discusses the central differences between the matching stage in this paper and the ones in Butters and other papers in the literature. In what follows, I outline the technical details of this setup.

To set the stage for the matching mechanism, consider a matching environment with M matches and V open vacancies. There are $L = U + s_e \cdot E + s_n \cdot N$ job seekers, where U, E and N are the number of workers in the unemployment, employment and non-participating states, respectively, and s_e (s_n) is the probability of search for an employed (non-participating) worker. Given some sufficiently small $\varepsilon > 0$, M, V and L correspond to natural numbers $n_M \in \mathbb{N}$, $n_V \in \mathbb{N}$ and $n_L \in \mathbb{N}$ in the following way: Let n_M, n_V and n_L be the smallest natural numbers satisfying $\varepsilon \cdot n_M \in B_\varepsilon(M), \varepsilon \cdot n_V \in B_\varepsilon(V)$, and $\varepsilon \cdot n_L \in B_\varepsilon(L)$. That is, the real numbers are a re-scaled stand-in for large natural numbers that have approximately the same ratio among each other as the original real numbers.[1] We are ultimately interested in the limit $\varepsilon \to 0$. However, this discrete representation allows us to think about matches, job seekers and vacancies in a discrete environment. There are two natural assumptions one could make about the relationship between matches, job seekers and vacancies. One could assume that matches are assigned to job seekers and vacancies in a way such that each vacancy and each job seeker receive at most one match. In this instance, if every match results in a job, each job seeker has a job finding probability of $\frac{n_M}{n_L} \to \frac{M}{L}$, which corresponds to the standard DMP matching mechanism used in most of the literature. I introduce an alternative assumption: Instead of being evenly distributed across job seekers and vacancies, I assume that meetings[2] are *randomly* assigned to a vacancy and a job seeker. Figure 10 illustrates the difference for the case $n_M = 4, n_V = 6, n_L = 5$.

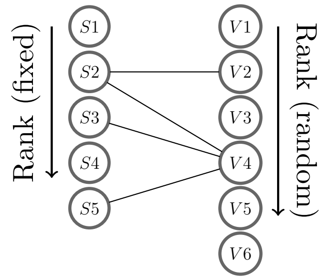
Under this assumption, any vacancy and any worker can end up with any number

[1] Strictly speaking, n_M, n_V and n_L are functions of ε and M, V and L, respectively, but for ease of notation this dependence is suppressed.

[2] In this environment, not all meetings can lead to successful matches. To streamline nomenclature, I shall refer to an encounter between worker and vacancy as a "meeting" or "encounter," and to a meeting that leads to a job as a "(successful) match."



(a) Standard matching



(b) Many-to-many

Figure 10: Illustration of the matching mechanism with $n_M = 4, n_V = 6, n_L = 5$

Notes: The left and right panel represent matching assumptions in the standard environment (left) and under many-to-many matching (right). Circles labeled "S" stand for job seekers, while circles labeled "V" stand for job vacancies. Black solid lines represent encounters.

of meetings, at most one of which can result in a successful match. I therefore call this matching process many-to-many matching, which, throughout the paper, I contrast with the one-to-one matching assumption that is standard in the search literature. For any vacancy j, denote by m_j the (random) number of meetings assigned to that vacancy and denote $q_\varepsilon = \frac{n_M}{n_V} \to q = \frac{M}{V}$. For vacancy j, the probability of getting $m_j = k$ meetings is

$$\begin{aligned}
P(m_j = k) &= \left(1 - \frac{1}{n_V}\right)^{n_M - k} \left(\frac{1}{n_V}\right)^k \binom{n_M}{k} \\
&= \left(1 - \frac{1}{n_V}\right)^{n_M - k} \left(\frac{1}{n_V}\right)^k \frac{n_M!}{k!(n_M - k)!} \\
&= \left(\left(1 - \frac{1}{n_V}\right)^{n_V}\right)^{q_\varepsilon} \left(1 - \frac{1}{n_V}\right)^{-k} q_\varepsilon^k \left(\prod_{j=0}^{k} \left(1 - \frac{j}{n_M}\right)\right) \frac{1}{k!} \\
&\xrightarrow{\varepsilon \to 0} e^{-q} q^k \frac{1}{k!}.
\end{aligned}$$

Hence, the limiting distribution of meeting frequency per vacancy is Poisson with parameter q. Employing an analogous argument, on the job seekers' side, we see that the number of meetings for a given searcher is Poisson-distributed with parameter λ. If we assume that each job seeker can take only one job and that vacancies can be filled with at most one worker, the potential multiplicity of meetings per searcher and per vacancy forces us to commit to a selection mechanism by which only a subset of meetings become successful matches. I will now describe this selection mechanism.

I assume that workers are ranked by their type i and vacancies are ranked randomly within each period.[3] Thus, every searcher can be characterized by their rank $p_L \in [0,1]$ within the distribution of job seekers. Likewise, every vacancy can be characterized by its (random) rank $p_V \in [0,1]$. Now, assume that at the start of a matching stage, all vacancies make offers in order of their rank p_V. Each vacancy starts out by making an offer to their encounter with the highest rank. If this job seeker has not previously received (and therefore accepted) an offer from a higher ranked vacancy, they accept the offer; otherwise, they reject. In case of rejection, the vacancy makes an offer to whoever has the second highest rank among their encounters. If this offer also gets rejected, they make an offer to the third-ranked encounter, and so on. In some cases, a vacancy might receive a positive number of meetings but not receive any successful match because the job seekers they meet all turn down the firm in favor of other offers.

To describe this process by means of example, consider Figure 10. Here, the highest-ranked vacancy, V1, has not encountered a job seeker and therefore stays unmatched (the same happens to V3, V5 and V6). V2 thus makes the first offer to S2, and S2 accepts. Next, V4 makes an offer to S2, but S2 is already matched and therefore rejects, so V4 makes an offer to S3, which is accepted. S5 ends the matching phase without a match, even though it has encountered V4, because S5 is outranked by S3. All remaining job seekers who have not encountered a vacancy also remain unmatched. This example illustrates why workers of higher rank generally have better chances of obtaining a match and provides intuition for why their matching probability is a smooth and increasing function of their rank i, which may vary with market tightness.

It is easy to show that this successive way of resolving encounters is conditionally Pareto optimal, even if workers are assumed to be indifferent between vacancies. Intuitively, every encountered worker with a rank higher than the one the firm matches with has been matched with another firm. To find a Pareto improvement, at least one firm has to be matched with a worker ranked higher than the one it was matched with in the original allocation. This means that another firm of higher rank will lose the match with this worker. To compensate this firm, a worker of even higher quality has to switch out of their old match into this firm, and so on. At some point, it will be impossible to compensate the firm, since it does not have a higher ranked encounter.

We can now determine the probability of a match given the rank of a particular worker.

[3] Since firms are assumed to be identical, a random ranking is natural in this setting

For this, define

$$
\begin{aligned}
f(p_L, p_V) &= P(p_L \text{ receives an offer from a vacancy of rank } p_V \text{ or higher}), \\
g(p_L, p_V) &= P(p_V \text{ receives no acceptance from a searcher of rank } p_L \text{ or higher}).
\end{aligned}
$$

For a given searcher, the number of meetings with firms greater than p_V is Poisson-distributed with parameter $\lambda(1 - p_V)$. Similarly, the number of meetings of the firm that feature a worker of rank greater than p_L is Poisson-distributed with parameter $q(1 - p_L)$. Let $\mu \sim \text{Pois}(\lambda(1 - p_V))$ be the number of meetings with vacancies ranked above p_V for a given searcher and $m \sim \text{Pois}(q(1 - p_L))$ be the number of meetings with job seekers ranked above p_L for a given vacancy. The probability that a searcher p_L receives an offer from a vacancy with rank p_V or higher is equal to the probability that at least one of the searcher's meetings higher than p_V gets rejected by all their meetings greater than p_L. The probability that a vacancy p_V receives no acceptance from a searcher of rank p_L or higher is equal to the probability that all the vacancy's meetings of rank above p_L get rejected by the worker (because that worker received an offer higher than p_V). Thus,

$$
\begin{aligned}
f(p_L, p_V) &= \mathbb{E}\left[1 - \left(1 - (1 - p_V)^{-1} \int_{p_V}^{1} g(p_L, \tilde{p}_V)\, d\tilde{p}_V\right)^{\mu}\right], \\
g(p_L, p_V) &= \mathbb{E}\left[\left((1 - p_L)^{-1} \int_{p_L}^{1} f(\tilde{p}_L, p_V)\, d\tilde{p}_L\right)^{m}\right],
\end{aligned}
$$

and therefore, using the fact that $\mathbb{E}\left[p^m\right] = \exp(\lambda(p - 1))$ for any number p and Poisson-distributed random variable $m \sim \text{Pois}(\lambda)$, we arrive at the following differential equation for f:

$$
1 - f(p_L, p_V) = \exp\left(-\lambda \int_{p_V}^{1} \exp\left(-q \int_{p_L}^{1} \left[1 - f(\tilde{p}_L, \tilde{p}_V)\right]\, d\tilde{p}_L\right) d\tilde{p}_V\right). \tag{8}
$$

It is straightforward to verify that this equation reduces to the edge case

$$
f(1, s) = 1 - \exp\big(\lambda(p_V - 1)\big)
$$

for $p_L = 1$. This is intuitive - a searcher who will be hired first by any firm will get a job offer better than p_V with exactly the probability of a meeting higher than p_V.

Differential equation (8) can be solved numerically, which yields solutions for f and g. The key object of interest is $f(p_L, 0)$, which describes the job finding probability for a given worker p_L. If we assume a Cobb-Douglas production function for meetings $M = aL^{\omega}V^{1-\omega}$, q and λ are related through the meeting efficiency parameter a by the formula

$$\lambda_t = a^{\frac{1}{\omega}} q_t^{\frac{\omega-1}{\omega}}. \tag{9}$$

This implies that for a given meeting efficiency a, it is possible to determine the job finding probability as a function of the rank within the distribution of job seekers using a single scalar, the number of meetings per searcher. The meeting efficiency a plays an important role for the relationship between a searcher's job finding rate and their rank. For larger values of a, the slope of the job finding curve will be steeper. Figure 11 shows how, when the average job finding probability is held constant, the job finding probability as a function of the searcher rank varies with the chosen value of the meeting efficiency a. Intuitively, as a goes to zero, there are far more vacancies than meetings (recall that we are

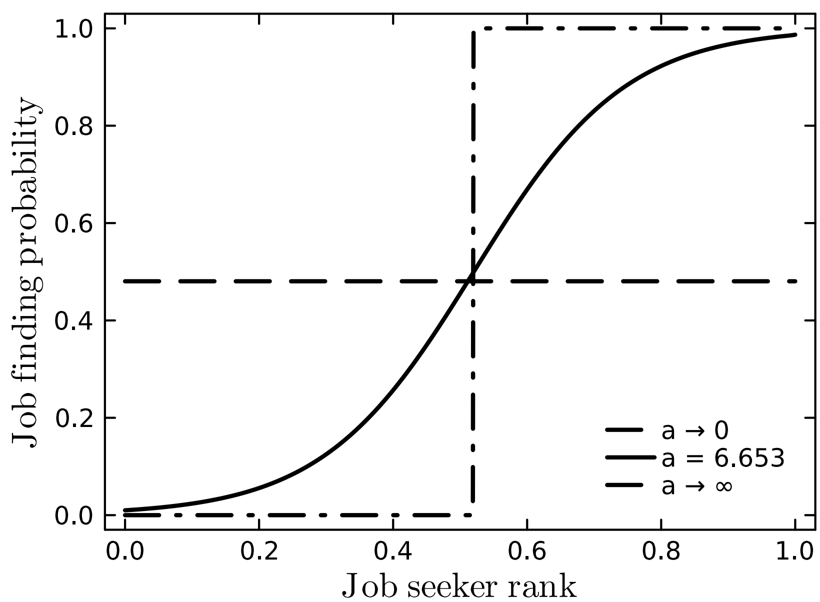


Figure 11: Job finding probability by rank for different values of a

Notes: Job finding probability in a given period as a function of the worker's rank in the job seeker distribution of a given period. Each line corresponds to one of three scenarios with different parameter values for the meeting efficiency parameter a and the same average job finding probability of job seekers. The middle value of $a = 6.653$ corresponds to the value chosen in the baseline calibration.

holding the average job finding rate constant). Therefore, meetings with a given vacancy are virtually guaranteed to be the vacancy's only contact, meaning that the probability of a match approximately equals the probability of a meeting, which is the same for all job seekers. As a goes to infinity, all vacancies are connected to all job seekers, meaning that vacancies can always find a match until there are either no vacancies or no job seekers left. Holding the average job finding rate constant, this corresponds to the situation where all vacancies end up with a match, but only the highest ranked job seekers match successfully. The value for a used in the baseline calibration introduces a relationship between rank and job finding probability between these two extreme cases.

Figure 12 illustrates the relationship between λ and the job finding rate for job seekers of various ranks, given the value of a used in the 2009-2020 calibration of the model. The figure shows that slack markets significantly lower the matching rates of workers at the bottom of the job seeker distribution but leave the matching rates of workers at the top mostly unaffected. The reason is that vacancies receive more applicants in slack markets, which allows firms to be more selective about whom they hire. This crowds out the lower end of the job seeker distribution in favor of higher ranked candidates, and workers at the bottom remain unmatched more often.

One way to think about this matching framework is to view the offer and response mechanism as a process during which some of the meetings created by the meeting function get destroyed while others turn into successful matches. Let $\tilde{\lambda}(p_L) = f(p_L, 0)$ denote the job finding rate - that is, the number of successful matches per searcher at rank p_L. Now, define

$$\sigma(p_L) = \frac{\tilde{\lambda}(p_L)}{\lambda} = \frac{f(p_L, 0)}{\lambda}. \tag{10}$$

Here, $\sigma(p_L)$ can be called the "survival function", and denotes the ratio of successful matches to total meetings at rank p_L. This function is fully determined by the state of the economy captured by λ (or, equivalently, q). Using the concept of the survival function, we can see that the model nests standard one-to-one matching: It can be verified that setting $\sigma(p_L) \equiv 1$ delivers the standard equations in this case.

Denote by $J(p_L)$ the firm value of a match with a worker of rank p_L. Denote by $\bar{J}$ the average value of a meeting to the firm. For a firm with a random vacancy rank, the ex-ante density of successful match probability by rank is given by $\sigma(p_L)$, and the expected

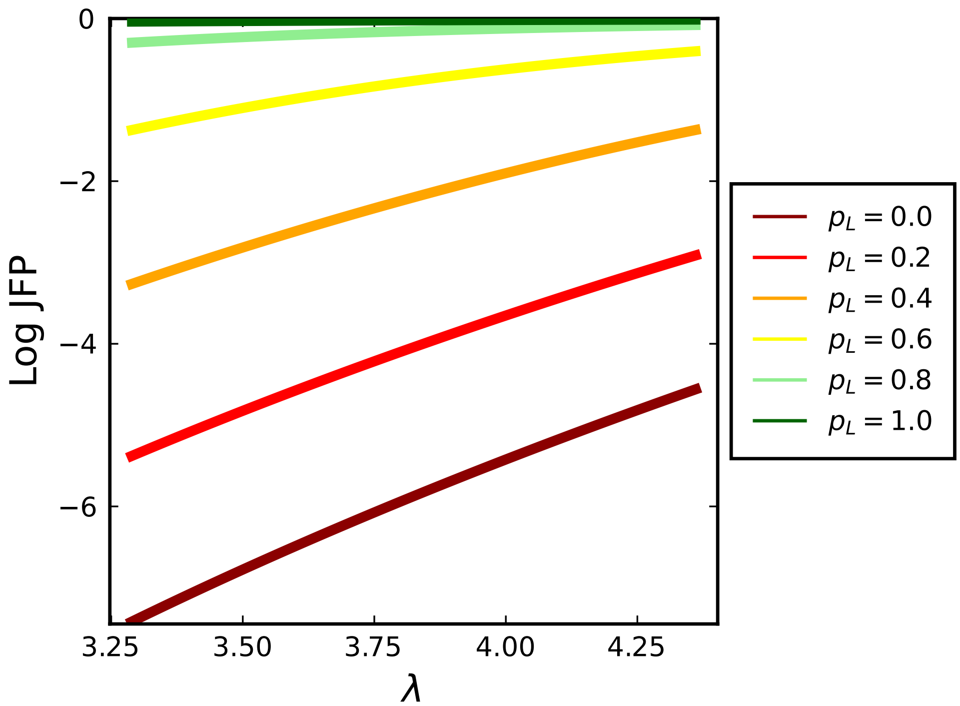


Figure 12: Job finding probability as a function of λ and p_L for $a = 6.65$

Notes: Model-predicted log job finding probability of a worker in a given quality quantile p_L of the job seeker distribution as a function of the aggregate encounter rate of workers λ, capturing aggregate market conditions.

average value of a meeting in period t is therefore given by

$$\bar{J}_t = \int_0^1 \sigma_t(p_L) J_t(p_L)\, dp_L. \tag{11}$$

In equilibrium, as in the DMP model, the cost of posting a vacancy has to equal the benefit. As the expected number of meetings per vacancy is q and the vacancy posting cost is κ, this leads to the well-known equilibrium condition

$$\kappa = q_t \bar{J}_t. \tag{12}$$

However, one feature of this equation sets it apart from the canonical DMP condition: Here, in contrast to the DMP model, $\bar{J}$ *directly* depends on q through $\sigma(.)$. For higher values of q, σ becomes more skewed towards higher values as firms receive more applications per vacancy and therefore become more selective. Note, however, that conditional on

$J(p_L)$, there is exactly one value of λ (or, equivalently, market tightness or q) that solves this equation. In this sense, there is still a direct mapping between the value of a meeting and the job finding rate, as in the DMP model. This simplifies immensely the computation of equilibrium because all transition probabilities into employment can be computed from the initial distribution and the vector $\{\lambda_\tau\}_{\tau=0}^{T}$, where T is some "final" period sufficiently far in the future. Hence, we will see that computing an equilibrium involves finding the sequence $\{\lambda_\tau\}_{\tau=0}^{T}$, which produces a sequence of distributions and transition probabilities that map into a series of meeting values $\bar{J}_t$, which, through equations 12 and 9, in turn correspond to $\{\lambda_\tau\}_{\tau=0}^{T}$.

3.3 Laws of motion and value functions

With this matching mechanism specified, the model is almost fully defined. We can write down transition probabilities from period to period by defining the following transition matrix:

$$\Theta_t(i) = \begin{pmatrix} (1-\delta_t^{nu})(1-s_n\sigma_t(p_L^t(i))\cdot\lambda_t) & \delta_t^{un} & \delta_t^{en}(i) \\ \delta_t^{nu} & (1-\delta_t^{un})(1-\sigma_t(p_L^t(i))\cdot\lambda_t) & (1-\delta_t^{en}(i))\delta_t^{eu}(i) \\ (1-\delta_t^{nu})s_n\sigma_t(p_L^t(i))\cdot\lambda_t & (1-\delta_t^{un})\sigma_t(p_L^t(i)\cdot\lambda_t & (1-\delta_t^{en}(i))(1-\delta_t^{eu}(i)) \end{pmatrix}.$$

This matrix enables us to conveniently write the value functions of a worker of type i in a non-participating, unemployed or employed state, denoted $V_t^N(i), V_t^U(i)$ and $V_t^E(i)$, respectively. Define $\mathcal{V}_t(i) = (V_t^N(i), V_t^U(i), V_t^E(i))'$. Then,

$$\mathcal{V}_t(i) = (b, b, w_t(i))' + \frac{1}{1+r}\left(\Theta_{t+1}^i\right)'\mathcal{V}_{t+1}(i). \tag{13}$$

Likewise, for the firm, denote by $J_t(i)$ the value of a successful match with worker i. This value is determined by the following Bellman equation:

$$\begin{aligned} J_t(i) =& y(i) - w_t(i) \\ &+ \frac{1}{1+r}\left[(1-\delta_{t+1}^{en}(i))(1-\delta_{t+1}^{eu}(i))(1-s_e\cdot\sigma_{t+1}(p_L^{t+1}(i))\cdot\lambda_{t+1})\right]J_{t+1}(i), \end{aligned} \tag{14}$$

where $p_L^t(i)$ is the rank of type i workers among job seekers in period t.[4] Finally, because of Nash bargaining, the wage $w_t(i)$ is set in every period such that

$$J_t(i) = (1-\beta)(J_t(i) + V_t^E(i) - \gamma_{t+1}(i)V_t^N(i) - (1-\gamma_{t+1}(i))V_t^U(i)), \tag{15}$$

where $\gamma_t = \delta_t^{en}(i)/(1-(1-\delta_t^{en}(i))(1-\delta_t^{eu}(i)))$ denotes the probability of separating into non-employment in period $t+1$ conditional on separating in period $t+1$.

Next, define $\mathcal{E}_t(i) = (N_t(i), U_t(i), E_t(i))'$. The law of motion of the type-state distribution is

$$\mathcal{E}_t(i) = \Theta_t(i)\mathcal{E}_{t-1}(i). \tag{16}$$

To determine the expected value of a meeting during a particular period, it is useful to change the integration measure of equation (11):

$$\begin{aligned} \bar{J}_t &= \left(\int_0^1 \sigma_t(\tilde{p}_L)\, d\tilde{p}_L\right) J_t \\ &= \overbrace{\left(\int_0^1 \sigma_t(\tilde{p}_L)\, d\tilde{p}_L\right)}^{(1)} \cdot \\ &\quad \int_0^1 \underbrace{\left(\frac{U_t^-(i) + s_n N_t^-(i) + s_e E_t^-(i)}{\int U_t^-(\iota) + s_n N_t^-(\iota) + s_e E_t^-(\iota)\, d\iota}\right)}_{(2)} \underbrace{\frac{\sigma_t(p_L^t(i))}{\int_0^1 \sigma_t(\tilde{p}_L)\, d\tilde{p}_L}}_{(3)} \underbrace{J_t(i)}_{(4)}\, di, \end{aligned} \tag{17}$$

where J_t denotes the average value of a match, $\bar{J}_t$ denotes the average value of a meeting, and $N_t^-(i) = (1-\delta_t^{nu})N_t(i)$, $U_t^-(i) = (1-\delta_t^{un})U_t(i)$ and $E_t^-(i) = (1-\delta_t^{en}(i))(1-\delta_t^{eu}(i))E_t(i)$ denote the number of non-participating, unemployed and employed workers of type i after the separation stage, respectively. Conditional on terms (2) and (4) in this equation, there is exactly one tuple $(\lambda_t, q_t, \sigma_t)$ (jointly determined by the matching function and equation (10)) that solves the free entry condition (equation (12)). That is, conditional on the composition of unemployed, employed and non-participating job seekers by type, and conditional on the value each type generates to the firm, the model endogenously generates the job finding rate conditional on type and a corresponding tightness.

[4]Note that this rank is not equal to i, because i denotes the worker's rank in the overall population, which is distinct from the population of job seekers.

We can now define an equilibrium.

Equilibrium: *Given an initial employment distribution* $\{N_0(i), U_0(i), E_0(i)\}_{i\in I}$ *and a path of separation and non-employment transition probabilities* $\{\delta_t^{eu}, \delta_t^{en}, \delta_t^{un}, \delta_t^{un}\}_{t=1}^{\infty}$*, an equilibrium is a sequence of type-state distributions* $\{\{N_t(i), U_t(i), E_t(i)\}_{i\in I}\}_{t=1}^{\infty}$*, worker meeting rates* $\{\lambda_t\}_{t=1}^{\infty}$*, firm meeting rates* $\{q_t\}_{t=1}^{\infty}$*, survival functions* $\{\sigma_t\}_{t=1}^{\infty}$*, worker values* $\{V_t^N(i), V_t^U(i), V_t^E(i)\}_{t=1}^{\infty}$*, firm meeting values* $\{\bar{J}_t\}_{t=1}^{\infty}$ *and wages* $\{w_t(i)\}_{t=1}^{\infty}$ *such that equations* (9)*,* (10)*,* (12)*,* (13)*,* (14)*,* (15)*,* (16)*, and* (17) *are satisfied and* $\sigma_t()$ *corresponds to the function* $f()$ *that solves equation* (8) *given* (λ_t, q_t)*.*

Equation (17) provides a useful decomposition of $\bar{J}_t$. First, a meeting can create value for the firm only when it turns into a match. Therefore, term (1) describes the average probability of a meeting becoming a successful match. The remaining terms (2, 3, 4) decompose the value of a successful match. Term (2) captures a composition effect: As the probability of search varies by employment state and different worker types have a different distribution over employment states, not all worker types are equally likely to be encountered by the firm, even though their distribution in the population is uniform by assumption. The value of a match therefore has to be adjusted for the likelihood of encountering particular worker types. Next, when there are several meetings per vacancy, firms will be able to choose the highest quality worker who accepts their offer. This skews the likelihood that a meeting will become a successful match towards higher searcher quantiles. This also has to be accounted for when calculating the value of a match which is done by including term (3), the "selection effect." Term (4) captures the value of a match conditional on worker type. I call fluctuations in this term the "direct effect."

In the model, the composition effect is at the heart of the persistently low UE rate during the early recovery. As unemployed workers are more likely to search than their non-participating or employed peers, the quality of the pool of job seekers is affected by the relative shares of different worker types in each employment state. As the recovery progresses, low job finding rates during the recession and in the beginning of the recovery preclude lower quality workers from matching and thus skew the pool of job seekers towards this group. This is captured by the composition effect. The composition effect plays a large role in making encounters less lucrative for firms, which depresses the number of vacancies. As a result, the job finding rate remains low, and the expected number of encounters per match stays high, which means that through selection, the composition effect reinforces itself.

The selection margin, on the other hand, plays an important role in generating differential recovery dynamics for the employed and the unemployed. In equilibrium, when the value of a job decreases, this term endogenously rises to offset this decrease because more selective firms will be able to hire workers from higher up in the quality distribution. In the beginning of a recovery, firms will cut back vacancies until the selection effect makes the benefit of posting a vacancy equal to its cost. This leads to more selective hiring, skewing the odds of successful job search more towards high quality workers, who are more likely to be employed.

4 Calibration

The parameters of the model can be divided into three different blocks. First, there are 3 aggregate scalar parameters, which I hold constant across recoveries: The meeting function coefficient on job seekers, ω, the bargaining weight of the worker, β, and the interest rate r. Even though it can be argued that all of these parameters may plausibly have changed across the recovery episodes in question, neither of the parameter choices considered has a strong impact on the results, for reasons explained below. Furthermore, neither parameter can be easily read off the data, which means that additional assumptions and more sophisticated estimation methods would need to be employed to recover them for each episode. To keep things simple, I opt to hold them constant over time instead. Table 2 summarizes their values and targets.

Consider ω first. In the many-to-many matching model, not all meetings translate into matches, and the average match probability conditional on a meeting depends on market tightness. Therefore, traditional estimates of the elasticity of the matching function with respect to job seekers (often assumed to be the unemployed only) estimate an object that is not identical to ω. However, the choice of ω is not critical to the results. The reason is that under high values of a, such as the one I calibrate, the number of successful matches is approximately equal to the minimum of job seekers and vacancies. Intuitively, a very large number of meetings means that in the limit, every job seeker encounters every vacancy, so that meetings turn into matches until there are either no remaining vacancies or no remaining job seekers. Mathematically, the effective matching function becomes $\tilde{M} = \min\{L, V\}$ for $a \to \infty$, regardless of ω. Thus, the choice of ω has limited influence on the results. Regardless, I inform ω using values typically found in the literature. That is, I follow the mean estimate in the survey of Petrongolo and Pissarides (2001) in setting $\omega = 0.4$. All results continue to hold for alternative choices of $\omega = 0.2$ or $\omega = 0.6$.

Next, consider β. Targets for the worker bargaining parameter are somewhat elusive in the literature. Since estimates of this parameter, as far as they exist, are based on standard one-to-one matching functions, these estimates are not precisely transferable to the setting many-to-many matching. Nonetheless, to keep things simple, I follow the common practice and choose a value corresponding to the meeting elasticity; that is, $\beta = \omega = 0.4$[5].

Turning to r, I choose an interest rate of 1% per annum. While this choice is somewhat ad hoc, it turns out that the interest rate also does not have a large effect in this model, as the discount factors relevant to the firm are mostly determined by separation rates which are orders of magnitude larger than any reasonable candidate values for the interest rate.[6]

Next, I turn to the set of parameters that are allowed to vary across recoveries. Consider first the parameters that govern the average transition probabilities of workers across states: $s_e, s_n, \delta^{eu}_{ss}, \delta^{en}_{ss}, \delta^{un}_{ss}, \delta^{nu}_{ss}$, which govern the average EE, NE, EU, EN, UN and NU probabilities, respectively. To discipline these parameters, I use data on transition probabilities from the basic monthly files of the Current Population Survey (CPS). I compute all gross transition flows from the CPS exactly as described in Shimer (2012). An exception to this is the series of job-to-job flows, which I take from Fujita, Moscarini and Postel-Vinay (2020) and adjust for seasonality as I do with the series from Shimer (2012).

For every recovery, I re-calibrate the relative search probabilities s_e, s_n as well as steady state transition parameters $\delta^{eu}_{ss}, \delta^{en}_{ss}, \delta^{un}_{ss}, \delta^{nu}_{ss}$ to match all respective aggregate period-to-period transition probabilities in steady state for every recovery, as detailed in Table 3. Target steady state period-to-period transition rates are defined as the average transition rate during the last 12 months of an observed recovery in the CPS. One exception to this is the EE transition probability for the first three recoveries in the sample, as the series from Fujita, Moscarini and Postel-Vinay (2020) extends back only to September 1995. I therefore match an EE transition probability of 2.83%, which is the average transition probability over the earliest 12 months observed in the EE series. I further calibrate κ so that the resulting steady state meeting rate λ_{ss} generates a steady state unemployment rate equal to the last unemployment rate observed during the recovery phase. Table 3 again summarizes parameter values and targets.

[5]This practice is commonly motivated with the Hosios condition. The Hosios condition is derived in a model environment with one-to-one matching and therefore does not apply in the present setting.

[6]This is largely a consequence of constant payoffs over the course of a job in this model. Back-loading of firm payoffs increases the dependency of job creation incentives on the interest rate. See, e.g., Hall (2017).

Finally, I turn to the parameters that govern cross-sectional worker heterogeneity. Following the empirical exercise in Section 2, one can again use the NLSY to discipline this heterogeneity. To do so, I use the classification of each worker according to their LJR centile q_{100} or their LJR quintile q_5.

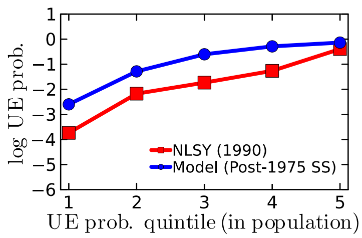
Consider first the worker's value of leisure b. For simplicity, I assume that $b(i) = b$; that is, there is no unobserved heterogeneity across worker types.[7] I set b equal to the minimum of the lowest observed steady state wage of a worker type in each episode, a value that corresponds to 70-76% of the average wage in each steady state.

Next, consider the meeting efficiency parameter a. This parameter governs the relative job finding rates of workers at the top versus the bottom of the rank distribution, as illustrated by Figure 11. Therefore, for each individual recovery episode, I set a to match the ratio of the average job finding probability of workers in the third quintile ($q_5 = 3$) relative to that of workers in the first quintile ($q_5 = 1$) in the last observed year of the recovery (e.g., 2019 for the 2009 recovery). For all episodes, I use the last recovery year, except for the 1975 and the 2020 recoveries, where I use data from 1990 and 2019 as respective calibration targets, as there is no data available for the actual reference years. Figure 13 shows a comparison between all quintiles of individual average job finding rates in the respective NLSY steady states compared with their model counterparts. The model has some difficulty matching the relative job finding rates between the fifth quintile and the fourth quintile, which is larger in reality than in the model. In addition, the job finding rates in the model are generally higher for each group than those in the NLSY. This happens partly because turnover is lower in the NLSY than in the CPS (which determines the target job finding rate in the model) and declining over time (pointing to age effects). However, the overall slope of job finding rates with respect to the cross-sectional quintile is relatively well matched by the model.

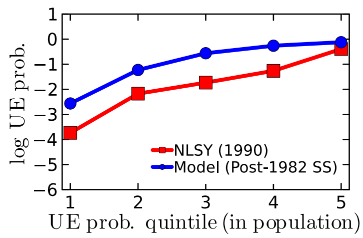
I now turn to the parameters for productivity and average likelihood of separation, which are specific to a worker's type i. To this end, I turn to the NLSY once more. First, I compute average EU and EN separation rates for each centile q_{100} and then run the regression

$$\log \mathrm{EU}_{q_{100,t}} = \eta_t + \xi_1 q_{100} + \xi_2 {q_{100}}^2 + \varepsilon^{EU}_{q_{100},t} \tag{18}$$

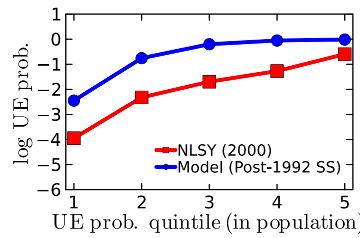
[7] Allowing for unrestricted heterogeneity here would require a reliable way to separately estimate $y(i)$ and $b(i)$ given wages, which is difficult.



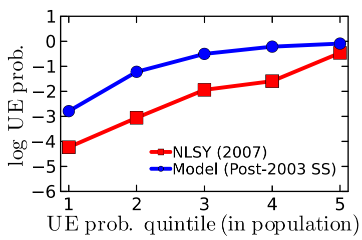
(a) 1975 recovery



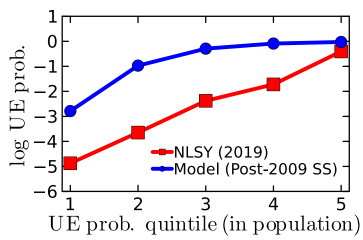
(b) 1982 recovery



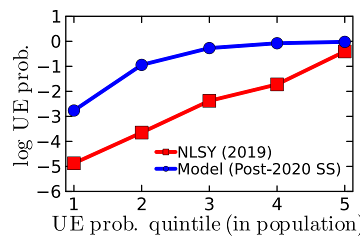
(c) 1992 recovery



(d) 2003 recovery



(e) 2009 recovery



(f) 2020 recovery

Figure 13: Cross-sectional job finding probabilities in model and data

Notes: Blue dots: Model-predicted log average job finding probability of workers in each population quintile for the indicated steady state, i.e., $\log \int_{\mathfrak{q}} \tilde{\lambda}_{ss}(p_L)\, dp_L$ for $\mathfrak{q} \in \{[0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1.0]\}$. Red dots: Empirical log average job finding probability for each quintile q_5 of workers across the distribution within the indicated year in the NLSY.

to project the worker-type-specific log EU rate onto a polynomial of degree two.[8] I proceed analogously to estimate worker-specific EN rates. Figure 14 shows the resulting estimates of $d^u(i) = \exp(\widehat{\log \mathrm{EU}}_{q_{100}(i),t})$ and $d^n(i) = \exp(\widehat{\log \mathrm{EN}}_{q_{100}(i),t})$. These functions are generally declining in worker type i, which indicates that higher quality workers tend to have more stable jobs.

Finally, I estimate worker-type-specific steady state wages $w_{ss}(i)$ by projecting each centile's average log wage in the last year of each recovery onto a second-order polynomial:

$$\log w_{q_{100}} = \xi_0^w + \xi_1^w q_{100} + \xi_2^w q_{100}^2 + \varepsilon_{q_{100}}^w.$$

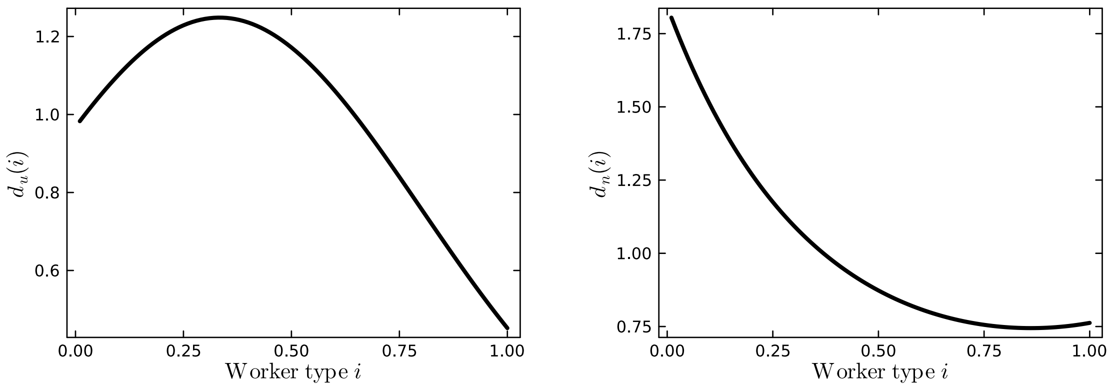
Then, I calibrate productivity levels $y(i)$ in every steady state to replicate the wage function $w_{ss}(i) = \exp(\widehat{\log w}_{q_{100}(i)})$ exactly.

It turns out that the wage function estimates from the NLSY for the 1990 steady state are somewhat atypical compared with those from other periods and feature an initially decreasing wage function. This somewhat odd empirical result echoes, in some ways, the observations from Section 2 where the data patterns of the youngest workers (until 1995) were shown to differ systematically from the data patterns of older workers (after 1995). It is not entirely clear why young workers' relative wages of low LJR individuals appear higher in 1990 than in later periods. In any case, taking this result seriously would make solving the model impossible for this period, as the selection rank would no longer plausibly be an increasing function of a worker's type. For these reasons, I opt to replace the wage function estimates from 1990 with those of 2000. Again, data are missing for the final steady state years of the 1975 and 2020 recoveries, which are replaced with data for the 1992 and 2009 recoveries (i.e., the 2000 and 2019 steady state wage functions), respectively. Figure 24 in the appendix summarizes the estimated productivity level as a function of i for all different episodes.

5 Results

In what follows, I use the calibrated model to simulate individual recoveries and benchmark it against a model with one-to-one matching. I apply the model to the most recent six labor market recoveries, defined as the period of time during which the unemployment

[8] Given the structural break observed in the data in Section 2, I exclude years before 1995 from the regression. I also exclude the lowest quantile $q_{100} = 1$, because it includes some individuals with characteristics that are systematically different from those of their low LJR peers and that tend to drive the results unless excluded. I do the same for the wage regression below.



(a) $d_u(i)$ (b) $d_n(i)$

Figure 14: Type-specific parameters

Notes: Estimated functional forms of $d_u(i)$ and $d_n(i)$ via equation (18) as described in the main text. Both functions are constructed from projections of EU and EN rates by worker centile q_{100} onto a second-order polynomial.

rate fell from its peak for a given recession to a stable trough (just before the subsequent recession). Table 5 summarizes these start and end points for the recoveries studied here. Each recovery is identified by its start year; that is, the year of peak unemployment.

I initialize a recovery as follows. Starting from steady state, five years before the peak of the unemployment rate, I choose a sequence of $\{\lambda_t, \delta_t^{eu}, \delta_t^{en}, \delta_t^{un}, \delta_t^{nu}, s_t^n\}$ so that the UE, EU, EN, UN, NU and NE transition probabilities from the data are exactly replicated. Note that during the run-up to the recovery (i.e., during the recession itself), I allow the probability of search for a non-participating worker to vary in order to enable precise matching of all transition rates simultaneously. This ensures that the stocks of non-participating, unemployed and employed workers track the empirically observed stocks closely. In period 0, before the first simulation period, I multiply δ_t^{eu} and δ_t^{en} by a factor that guarantees a perfect match of the unemployment rate in this period. I then start the simulation, fixing s_n at its calibrated value for the remainder of the simulation. This means that from that point forward, the value of a job and all distributions move endogenously as determined by the model equilibrium. In particular, conditional on the initial distribution and the aggregate separation paths into unemployment and non-participation, there is no further input from the data into the model. Any similarity of the recovery between model and data is therefore generated endogenously; only separations and transitions among the two non-employment states are taken as given.

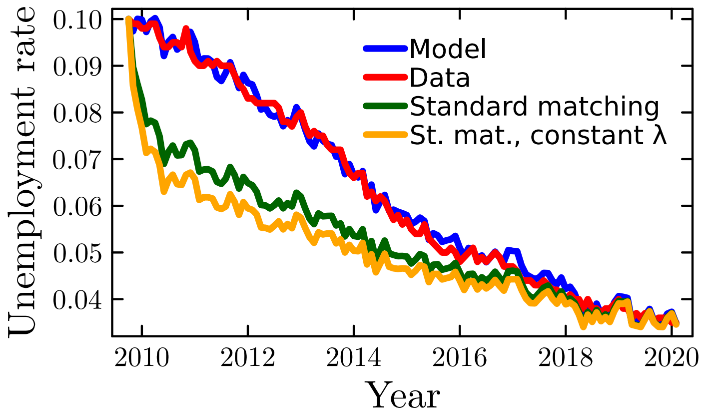


Figure 15: Unemployment: Data vs. baseline vs. standard matching (2009 rec.)

Notes: Red line: BLS unemployment rate during the recovery from the Great Recession. Blue line: Unemployment rate as predicted by the baseline model. Green line: Unemployment rate as predicted by the model with standard matching ($\sigma(p_L) \equiv 1$). Yellow line: Unemployment rate in the model with standard matching and fixed encounter rate ($\lambda = \lambda_{ss}$).

5.1 Unemployment and transition rates

Figure 15 shows the modeled recovery path along with the path observed in the data for the recovery following the Great Recession, which is the longest and arguably most well-known episode in the sample. As a reference, the figure also shows two other recovery paths generated by modifications of the model. The first modification is a (re-calibrated) version with the same initial state distribution and standard "DMP-style" one-to-one matching (i.e., $\sigma(p_L) \equiv 1$). The second is a version with DMP-style matching where the job finding rate λ is held constant at its steady state level. The first serves as a benchmark to show that a DMP-style framework cannot solve the recovery puzzle. The second illustrates the reason for this shortcoming: The one-to-one model does not generate a large enough drop in unemployed workers' job finding rates to allow the model to match the observed recovery. In the one-to-one matching counterfactual, unemployment remains at a level above its steady state for a while because exogenous separations increase the inflow into unemployment. In addition, increased separation rates and the initial composition effect cause the job finding rate to drop slightly. However, the drop is small, and thus the unemployment rate barely rises above the value of unemployment in the counterfactual with a constant job finding rate. Thus, over the course of the recovery, the unemployment rate implied by the one-to-one matching model quickly falls to a value well below its starting point.

In contrast, the many-to-many matching model generates an unemployment series that follows the series observed in the data very closely. The model delivers the correct recovery dynamics because it generates the correct average job finding rate for unemployed workers, as confirmed by Figure 16: The model generates a post-recession drop in the UE transition probability that only slowly adjusts upward as the recovery progresses.

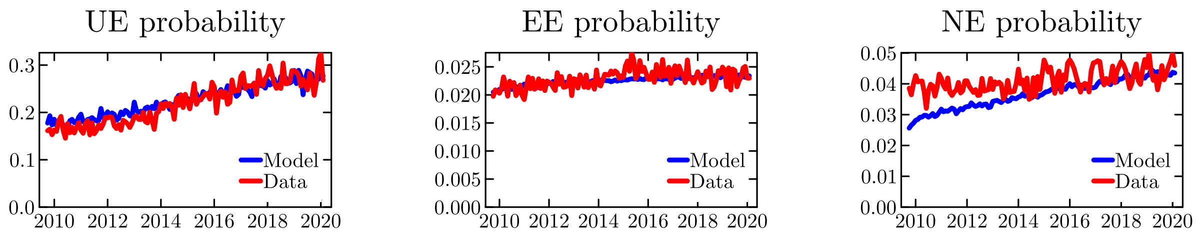


Figure 16: Transition probabilities in model and data (2009 recovery)

Notes: Red line: Gross worker flows in the CPS (UE, NE gross flows computed as in Shimer (2012), EE flows from Fujita, Moscarini and Postel-Vinay (2020)). Blue line: Worker flows implied by the baseline model.

Figure 16 also highlights another important feature of the model: As in the data, while the UE transition probability remains well below steady state level during the early recovery, the EE transition probability is much less depressed during this period. This feature is generated endogenously through the ranking assumption of the model: The employed are typically higher ranked workers who still tend to end up on top of their encountered vacancies' applicant pools even if market tightness is low and every vacancy receives on average more encounters than in steady state. The unemployed, however, tend to be lower ranked workers. If they encounter a vacancy, they are likely to be outranked by another worker. The post-recession environment amplifies this effect: During times of low market tightness, there are more applicants for a given position, and as a result, lower ranked workers are more likely to be outranked. This slows down the recovery of unemployment, because more selective firms shift their hiring towards the employed and the outflow from the stock of unemployed workers decreases. Finally, Figure 16 also shows that the model somewhat understates the NE transition rate during the early recovery.[9]

[9] There are at least two potential reasons for this: The first is that the model assumes the UN or NU transition probabilities to be independent of type, which mechanically generates similar compositions for both pools. In reality, there might be a composition of worker types in the pool of non-participating that tends to make this pool more successful. Secondly, the NE margin might be driven in part by a matching technology that lies outside the traditional random-search framework, such as recalls into old jobs after

5.2 The role of selection

To highlight the role that selection plays in this process, Figure 17 shows the counterfactual selection probabilities if every worker transitioned into employment with the average aggregate probability (i.e. if workers were selected randomly but vacancy creation and all other variables were held constant). The figure illustrates that selection depresses the

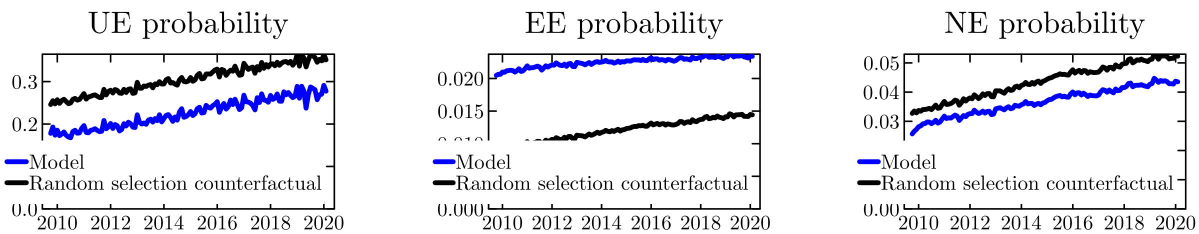


Figure 17: Actual and no-selection transition probabilities in the baseline model

Notes: Blue line: Worker flows implied by the baseline model. Black line: Counterfactual worker flows implied by the baseline model if workers were randomly selected.

job finding rate for the unemployed and elevates it for the employed. This pattern is particularly salient during the beginning of the recovery, where the (percentage) drop in the UE transition probability would be weaker without selection. The figure also shows that employed job seekers are more than twice as successful looking for a job as they would be if they did not enjoy a ranking advantage.

An alternative way to assess the role of selection is to recalibrate a model version in which selection is eliminated by randomly ranking all workers each period.[10] Importantly, this is not identical to the standard matching mechanism, as the model retains the feature of multiple encounters and therefore also the properties of the many-to-many meeting function, aside from selection. Figure 18 shows the implied transition rates for this model. Again, the UE rate suffers a far less severe drop early during the recovery in the no selection model compared to the baseline model and the data. The EE rate drops more than both the baseline model and the data. The NE rate likewise is slightly more stable than in the baseline model, although the difference is minor.

Overall, these figures show that selection plays an important role in amplifying the effect of slack markets on the exit rate out of unemployment. However, to fully understand

taking time off. However, it is clear that this mismatch between model and data does not adversely affect the model fit for the unemployment rate.

[10] To facilitate comparison, a is set at the same level as in the baseline.

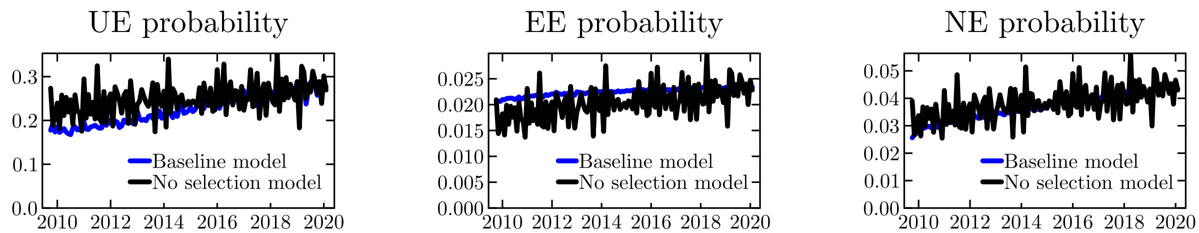


Figure 18: Transition probabilities in the baseline and no-selection models

Notes: Blue line: Worker flows implied by the baseline model. Black line: Worker flows implied by the model with random worker selection.

the recovery dynamics resulting from the model, we now need to establish why market tightness is low in the aftermath of a recession and does not quickly revert back to steady state after the shock has disappeared, as it does in the baseline DMP model.

5.3 Forces depressing job creation

To analyze the forces responsible for keeping market tightness persistently low, recall that one can map $\bar{J}_t$ and λ into one another per equations (9) and (12), which jointly yield

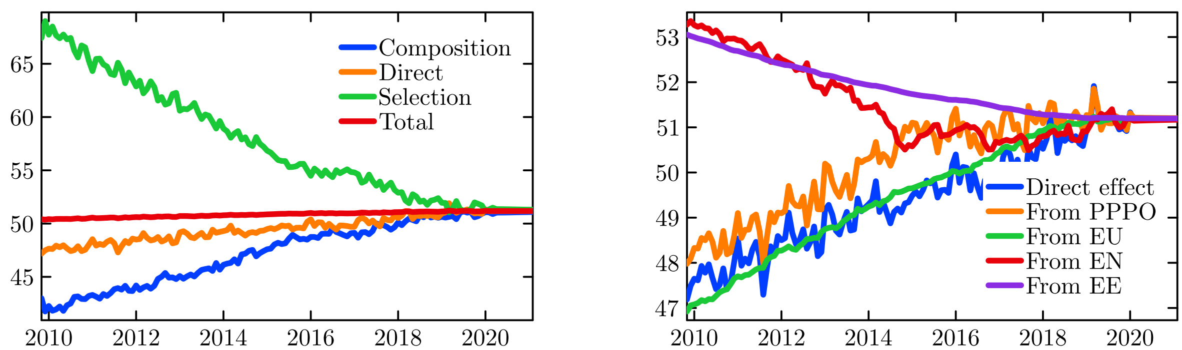
$$\lambda_t = a^{\frac{1}{\omega}} \left(\frac{\bar{J}_t}{\kappa} \right)^{\frac{1-\omega}{\omega}} = a^{\frac{1}{\omega}} \left(\frac{S_t J_t}{\kappa} \right)^{\frac{1-\omega}{\omega}} ,$$

where $S_t = \int_0^1 \sigma_t(\tilde{p}_L)\, d\tilde{p}_L$. Recall that S_t is solely determined by λ_t and therefore can be thought of as the output of a function that takes λ as its argument. In a slight abuse of notation, we can call this function S and thus write $S_t = S(\lambda_t)$, which enables us to rewrite the above equation as

$$\lambda_t \cdot S(\lambda_t)^{\frac{\omega-1}{\omega}} = a^{\frac{1}{\omega}} \left(\frac{J_t}{\kappa} \right)^{\frac{1-\omega}{\omega}} .$$

Next, it can be shown numerically that the left hand side expression is an increasing function of λ_t in the region of interest (that is, for all values of λ_t observed during any recovery). Thus, the number of meetings per job seeker is a monotone transformation of the firm value of a successful match. Therefore, a decomposition of J_t is informative about the fundamentals that determine the state of the labor market, captured by λ_t.

To understand the drivers of λ_t, we can use equation (17) to decompose the average value of a match J_t into changes in the three terms (2), (3) and (4). Concretely, I decompose J_t by holding two out of the three terms at their steady state level while letting the remaining one follow the equilibrium path from the model simulation. Figure 19a shows the result of this decomposition for the Great Recession. The figure shows that the



(a) Decomposition of J_t (2009 rec.)

(b) Direct effect decomp. (2009 rec.)

Figure 19: Decomposition of J_t and the direct effect

Notes: Left panel: Decomposition of J_t according to equation (17). Each line corresponds to the counterfactual path of J_t if only one of the three terms in equation (17) is allowed to vary as predicted by the model. Right panel: Decomposition of the direct effect (corresponding to term (4) in equation (17)). This decomposition relies on equation (14) and is described in the main text.

composition effect plays the most significant role in depressing the value of a match. The composition of job seekers is skewed towards worse workers initially, which comes from the low market tightness during the recession. This composition effect is then propagated as high quality workers continue to quickly transition into employment while lower quality workers find it difficult to find jobs under the slack labor market conditions present during the early phase of the recovery. Only over time, this effect becomes less important, and market conditions relax to the point where the dynamics reach steady state.

This finding is consistent with the evidence presented in Section 2 and speaks to a long and active debate in the literature on whether composition-based explanations for persistent unemployment fluctuations are consistent with the data. Some authors (Barnichon and Figura, 2015; Gregory, Menzio and Wiczer, 2024) have argued that the pool of unemployed workers becomes worse in the immediate aftermath of a recession, and this idea is built into many models of unemployment fluctuations with worker heterogeneity, such as those of Pries (2008), Ravenna and Walsh (2012) or Ferraro (2018). Barnichon and Figura (2015) provide empirical evidence that in the aftermath of the Great Recession, the pool of unemployed workers shifted towards workers with lower job finding rates, which is consistent with the finding in this paper. For the same episode, Gregory, Menzio and Wiczer (2024) argue that the pool of unemployed shifted towards workers with high separation rates. On the other hand, Mueller (2017) shows that in the CPS, the composition of unemployed workers shifts towards higher wage workers in recessions. He thus argues that the composition of job seekers is countercyclical.

As discussed in Section 2, my setting differs in two respects from these papers. For one, in addition to the unemployed, both the employed and the non-participating search for jobs, and the composition of the pool of all job seekers is what matters for vacancy posting incentives. Secondly, from the perspective of firms, not all workers are equally relevant. Firms weigh each worker's value by the likelihood that they will hire this worker. They do so in order to determine the aggregate average value per match, in line with equation (17). To see this formally, recall that the model-consistent notion of a composition effect is given by

$$\text{CompositionEffect}_t = \int_0^1 \text{Prevalence}_t(i) \cdot \text{SelectionWeight}(i) \cdot \text{Quality}(i)\, di, \tag{19}$$

where

$$\begin{aligned}
\text{Prevalence}_t(i) &= \left(\frac{U_t^-(i) + s_n N_t^-(i) + s_e E_t^-(i)}{\int U_t^-(\iota) + s_n N_t^-(\iota) + s_e E_t^-(\iota)\, d\iota} \right), \\
\text{SelectionWeight}(i) &= \frac{\sigma_{ss}(p_L^{ss}(i))}{\int_0^1 \sigma_{ss}(\tilde{p}_L)\, d\tilde{p}_L}, \\
\text{Quality}(i) &= J_{ss}(i).
\end{aligned}$$

This expression captures a notion of the composition effect that is consistent with the model. However, this measure can be adapted to yield alternative measures of worker quality that are not directly relevant to the model but yield intuitive measures of average job seeker quality that may matter for firm hiring in alternative model environments or yield empirically identifiable proxies for the expression above.

One alternative measure of composition effects can be constructed by setting $\text{Quality}(i) = i$ - that is, measuring quality as the average worker rank i instead of the firm value of a job $J_{ss}(i)$. This yields a notion of worker quality that is similar to the model expression but also can be empirically identified using LJR quantiles. This notion of average worker quality is used in the quality measures in Section 2. Secondly, one can decide to focus on the unemployed only and set $\text{Prevalence}_t(i) = U_t^-(i)/(\int U_t^-(\iota)\, d\iota)$, which yields a measure of average quality of unemployed workers. Finally, one can drop the SelectionWeight term ($\text{SelectionWeight}(i) = 1$) to get "unweighted" estimates of worker quality that correspond to simple averages of worker quality in the pool of job seekers.

Motivated by the above measures and echoing Section 2, I plot several different measures of the average quality of job seekers in the model in Figure 20. Using the figure, I show that accounting for search of non-participating and employed workers and weighting them by their selection probability, as is consistent with the model-implied measure, increases the procyclicality of job seeker quality.

The figure's two panels show two distinct measures of worker quality: The firm's value a job $J_{ss}(i)$, which is the "true" model-consistent notion of worker quality (left panel), and the worker's rank i, which is the empirically measurable notion of worker quality, as implemented in Section 2 (right panel). Since $J_t(i)$ increases near linearly in i, the two notions are near proportional.

The gray line traces out the unweighted average quality of the pool of unemployed workers ($\text{Prevalence}_t(i) = U_t^-(i)/(\int U_t^-(\iota)\, d\iota)$, $\text{SelectionWeight}(i) = 1$) in both panels. This compositional measure is comparable with measures used in other papers in the literature, which have often focused on the unemployed only. It is very mildly procyclical but almost flat, in line with the observations in Section 2. The black line instead shows the average quality of job seekers when each worker is weighted by their likelihood of selection ($\text{Prevalence}_t(i) = U_t^-(i)/(\int U_t^-(\iota)\, d\iota)$, $\text{SelectionWeight}(i) = \sigma_{ss}(p_L^{ss}(i))/(\int_0^1 \sigma_{ss}(\tilde{p}_L)\, d\tilde{p}_L)$). This weighting amplifies the cyclicality of the composition of the pool of unemployed workers. The same pattern can be observed when one focuses on the pool of all job seekers ($\text{Prevalence}_t(i)$ as in the baseline measure). Here, weighting workers by their selection probability again significantly increases the procyclicality of job seeker quality. Both these findings are in line with the evidence presented in Section 2.

Thus, while the model does imply procyclical composition effects, the average quality of workers in unemployment is nearly acyclical. Figures 30-34 in the appendix show that this pattern can also be documented for the other recoveries in my sample.

Next, I briefly turn to the direct effect. It can be further decomposed into the effects coming from the different components of the discount factor as well as the effect that comes from the per-period payoff (PPPO) for the firm through the interplay of market conditions and wages. To do this, I perform a decomposition of equation (14) by the same principle as applied above: I plot J_t while holding all distributions constant at their steady state and holding constant all but one of $\delta_t^{en}(i), \delta_t^{eu}(i)$, $\sigma_t(p_L(i))\lambda_t$, and $w_t(i)$. Figure 19b plots this decomposition of the direct effect. It is clear that the main effect depressing the direct effect is the transition probability from employment into unemployment, which remains above steady state for a long time into the recovery. The transition probability from employment into non-participation is depressed during the recovery, increasing the value of a match. The transition rate into employment is also depressed as the aggregate job finding probability falls. This, too, works towards increasing the value of a job. Finally, wages have a moderate downward effect on the job value because reduced poaching rates in the early recovery moderately raise the value of a job and the worker extracts some of the resulting surplus by negotiating a higher wage.

I now turn to other recoveries and conduct the same exercise. For each recovery, I re-calibrate the steady state value of all δs and κ to match new steady states. Figure 21 shows the unemployment fit for these other recoveries. The fit of the simulated unemployment series compared with the data is excellent for three out of the four recoveries (1975, 1992 and 2003) and good for the fourth (1982). The simulation of the 1982 recovery is somewhat of an outlier in that the unemployment rate plateaus at a level that is slightly too high towards the first half of the recovery. Aside from this, the many-to-many matching model is able to endogenously generate the exact shape of recovery observed after every of the last five major pre-COVID recessions in the US. The model replicates the series both for recoveries with a large direct contribution from the separation margin (such as 1975 and 2003) and for recoveries in which this direct effect is not nearly sufficient to explain the elevated unemployment rate (such as 1992 and 2009).

The model is also successful in capturing some of the quality difference between the pool of employed and unemployed job seekers that the literature has documented in the data. Computing the wage premium for hires out of employment relative to hires out of unemployment, I find that in the steady state at the end of the 2009 recovery, the average hire from employment goes on to earn a wage that is 9 log points higher than the average wage of a hire out of unemployment. In the data, Faberman et al. (2017) estimate this premium to be 36 log points, of which 17 log points can be explained with observable characteristics.

5.4 Vacancies

We now turn to the model's implications for vacancies. As discussed in Section 4, the relatively high calibrated value for the meeting efficiency a means that matches track vacancies closely, whereas vacancies reported in JOLTS tend to be more volatile than hires over the cycle. This means that, essentially by construction, the model cannot replicate the empirical volatility of vacancies in JOLTS. However, JOLTS vacancies are measured only once at the end of the month, which generates a disconnect between vacancies posted over the course of the month and their stock at the end of the month when their average duration fluctuates. If one sees the empirical analogue of vacancies in our discrete time model as the number of postings per month, then an empirical measure of this variable would be the number of JOLTS vacancies divided by their average vacancy duration. Doing so using data from Davis, Faberman and Haltiwanger (2013) yields estimates of the number of vacancies posted within a month that co-move closely with the number of matches.[11]

Figure 22 shows the Beveridge curves implied by the model. The Beveridge curves exhibit slopes smaller than those typically found in the data, which is unsurprising given the discussion above. However, the model does replicate the inward shift of the Beveridge curve over more recent recoveries noted by Elsby, Michaels and Ratner (2015).

5.5 The COVID recovery

The model is able to reproduce the surprising regularity in the behavior of unemployment rates during the 1975, 1982, 1992, 2003 and 2009 recoveries. However, in more recent history, the US has experienced another recovery that falls a little outside the pattern established by its predecessors: the labor market readjustment after the COVID pandemic of 2020. In many respects, the COVID recovery is unusual. Transition dynamics in and out of unemployment were confounded by a large number of recalls into old jobs as demand in the US economy began to recover and supply restrictions were loosened. The unemployment rate jumped up to 14.7% but then adjusted much faster than in previous recoveries. Moreover, the 2020 recovery is the only one in the sample whose dynamics are driven primarily by movements in separations rather than the UE rate, as shown in Figure 2. This arguably poses an interesting challenge to the model: Can the selection-based model predict the dynamics that arose from the 2020 recovery?

Figure 23 shows the model-implied dynamics. Although standard matching and many-to-many matching perform equally well in the initial phase of the recovery, the baseline model gains the upper hand in the latter phase of the recovery. Notably, the two models

[11] Some authors have also emphasized the disconnect between recruiting effort and vacancy posting (see Gavazza, Mongey and Violante, 2018), which is another reason that the concept of a vacancy in this model does not necessarily correspond perfectly to vacancies as observed in JOLTS.

do not differ much in their implications for the unemployment rate, in contrast to their divergent predictions for other recoveries.

The second panel of Figure 23 reveals why the model is successful in capturing unemployment dynamics during the COVID recovery even though it is qualitatively different from previous episodes. It shows that the negative composition effects that slow down other recoveries are also operative during the COVID recovery, but the quick build-up of the COVID recovery leads to initial conditions that feature a positive initial composition of the pool of job seekers. This limits the negative feedback loop that slows down other recoveries and leads to a much more rapid adjustment of unemployment back to steady state.

6 Conclusion

This paper makes two contributions. First, it provides new evidence on compositional changes in the pool of job seekers and links it to differences in the cyclicality of job finding rates across groups. Motivated by these findings, it develops a matching model with worker heterogeneity and two-sided multiplicity of encounters that models worker selection in a tractable way. With selection, more productive workers return to employment more quickly than their less productive peers. Selection provides a rationale for the empirical finding that workers' job finding probabilities are more sensitive to the business cycle for workers at the lower end of the quality distribution than for those at the top. I then show that in a heterogeneous agent economy with selection, a feedback loop between composition effects and selection can generate the correct dynamics of the UE transition probability for the six most recent recoveries in US data. This feature allows the model to match each recovery's unemployment dynamics given the observed path of separations. The changing composition of the pool of job seekers and selection can therefore be considered a primary source of sluggish adjustment of the unemployment rate during these episodes.

The ranked matching mechanism opens up more avenues for future research. In particular, it links the efficiency of the meeting technology to inequality in labor market outcomes. While I have abstracted from changes in the meeting efficiency, a model along these lines could assess the effects of better matching technologies such as the rise of online application systems on inequality. I leave this question for future research.

In this paper, I take separation rates as given and therefore sidestep the perhaps more fundamental question of why they stay elevated for a long time after the initial recession shock has subsided. This is another question future research should address.

References

Andolfatto, David. 1996. "Business Cycles and Labor-Market Search." *American Economic Review*, 86(1): 112–132.

Barnichon, Regis, and Andrew Figura. 2015. "Labor Market Heterogeneity and the Aggregate Matching Function." *American Economic Journal: Macroeconomics*, 7(4): 222–249.

Barnichon, Regis, and Yanos Zylberberg. 2019. "Underemployment and the Trickle-Down of Unemployment." *American Economic Journal: Macroeconomics*, 11(2): 40–78.

Birinci, Serdar, Kurt See, and Shu Lin Wee. 2024. "Job Applications and Labour Market Flows." *Review of Economic Studies*, rdae064.

Blanchard, Olivier Jean, and Peter Diamond. 1994. "Ranking, Unemployment Duration, and Wages." *Review of Economic Studies*, 61(3): 417–434.

Bradley, Jake. 2022. "Worker-Firm Screening and the Business Cycle."

Butters, Gerard R. 1977. "Equilibrium Distributions of Sales and Advertising Prices." *Review of Economic Studies*, 44(3): 465–491.

Davis, Steven J, R Jason Faberman, and John C Haltiwanger. 2013. "The Establishment-Level Behavior of Vacancies and Hiring." *Quarterly Journal of Economics*, 128(2): 581–622.

Eeckhout, Jan, and Ilse Lindenlaub. 2019. "Unemployment Cycles." *American Economic Journal: Macroeconomics*, 11(4): 175–234.

Elsby, Michael WL, Ryan Michaels, and David Ratner. 2015. "The Beveridge Curve: A Survey." *Journal of Economic Literature*, 53(3): 571–630.

Engbom, Niklas. 2021. "Contagious Unemployment."

Faberman, R Jason, Andreas I Mueller, Ayşegül Şahin, and Giorgio Topa. 2017. "Job Search Behavior among the Employed and Non-Employed."

Ferraro, Domenico. 2018. "The Asymmetric Cyclical Behavior of the US Labor Market." *Review of Economic Dynamics*, 30: 145–162.

Ferraro, Domenico. 2023. "Fast Rises, Slow Declines: Asymmetric Unemployment Dynamics with Matching Frictions." *Journal of Money, Credit and Banking*, 55(2-3): 349–378.

Fujita, Shigeru, Giuseppe Moscarini, and Fabien Postel-Vinay. 2020. "Measuring Employer-to-Employer Reallocation."

Gavazza, Alessandro, Simon Mongey, and Giovanni L Violante. 2018. "Aggregate Recruiting Intensity." *American Economic Review*, 108(8): 2088–2127.

Gregory, Victoria, Guido Menzio, and David Wiczer. 2024. "The Alpha Beta Gamma of the Labor Market." *Journal of Monetary Economics*, 103695.

Hagedorn, Marcus, and Iourii Manovskii. 2008. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." *American Economic Review*, 98(4): 1692–1706.

Hall, Robert E. 2017. "High Discounts and High Unemployment." *American Economic Review*, 107(2): 305–330.

Hall, Robert E, and Marianna Kudlyak. 2022. "Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?" *NBER Macroeconomics Annual*, 36: 1–55.

Menzio, Guido, and Shouyong Shi. 2011. "Efficient Search on the Job and the Business Cycle." *Journal of Political Economy*, 119(3): 468–510.

Mercan, Yusuf, Benjamin Schoefer, and Petr Sedláček. 2021. "A Congestion Theory of Unemployment Fluctuations."

Merz, Monika. 1995. "Search in the Labor Market and the Real Business Cycle." *Journal of Monetary Economics*, 36(2): 269–300.

Mueller, Andreas I. 2017. "Separations, Sorting, and Cyclical Unemployment." *American Economic Review*, 107(7): 2081–2107.

Oh, Ryungha. 2024. "Spatial Sorting of Workers and Firms." Technical report.

Petrongolo, Barbara, and Christopher A Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature*, 39(2): 390–431.

Pries, Michael J. 2008. "Worker Heterogeneity and Labor Market Volatility in Matching Models." *Review of Economic Dynamics*, 11(3): 664–678.

Ravenna, Federico, and Carl E Walsh. 2012. "Screening and Labor Market Flows in a Model with Heterogeneous Workers." *Journal of Money, Credit and Banking*, 44(S2): 31–71.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95(1): 25–49.

Shimer, Robert. 2006. "On-the-Job Search and Strategic Bargaining." *European Economic Review*, 50(4): 811–830.

Shimer, Robert. 2012. "Reassessing the Ins and Outs of Unemployment." *Review of Economic Dynamics*, 15(2): 127–148.

Wolthoff, Ronald. 2018. "Applications and Interviews: Firms' Recruiting Decisions in a Frictional Labour Market." *Review of Economic Studies*, 85(2): 1314–1351.

A Tables

Table 1: Sensitivity of transition probabilities across groups

UE Prob. Quantile (q_5)	1st	2nd	3rd	4th	5th
Coefficient $\beta_1^{q_5}$	-1.23	-0.49	-0.31	-0.07	-0.14
	(0.27)	(0.19)	(0.12)	(0.10)	(0.09)
Coefficient $\delta_1^{q_5}$	0.71	0.72	0.57	0.53	0.47
	(0.39)	(0.18)	(0.20)	(0.14)	(0.22)

Notes: Regression coefficients of regressions (4) and (5), conducted for each individual LJR quintile q_5. Robust standard errors are shown in parentheses. The regression is conducted at a yearly level based on all individuals in the NLSY sample over the entire data period (1982-2019).

Table 2: Externally calibrated values

Single parameters		
Parameter	Value	Explanation
ω	0.4	Petrongolo and Pissarides (2001)
β	0.4	$= \omega$
r	0.01 p.a.	

Distributional parameters	
Parameter	Target
$d^u(i)$	relative EU probability by worker type
$d^n(i)$	relative EN probability by worker type
$y(i)$	$w_{ss}(i)$ (average wage by worker type)

Notes: Scalar parameters are set externally; distributional parameters are estimated outside the model. Scalar parameter values are displayed in the "value" column. The functions $y(i)$, $d_u(i)$ and $d_n(i)$ are estimated as described in the main text, and $d_u(i)$ and $d_n(i)$ are shown in Figure 14, while $y(i)$ is shown in Figure 24.

Table 3: Internally calibrated values

Target value							
Parameter	Target	1975	1982	1992	2003	2009	2020
κ	ur_{ss}	0.06	0.055	0.039	0.047	0.035	0.035
s_e	EE	0.0283	0.0283	0.0283	0.0241	0.0233	0.024
s_n	NE	0.0495	0.0489	0.0498	0.0473	0.0438	0.044
δ^{eu}_{ss}	EU	0.0146	0.0145	0.0113	0.0116	0.0088	0.0093
δ^{en}_{ss}	EN	0.0337	0.0285	0.0286	0.0294	0.031	0.0325
δ^{nu}_{ss}	NU	0.0244	0.0228	0.0209	0.0208	0.0155	0.0149
δ^{un}_{ss}	UN	0.229	0.213	0.253	0.245	0.258	0.264
Parameter value							
Parameter	Target	1975	1982	1992	2003	2009	2020
κ	ur_{ss}	33.56	33.23	34.9	41.72	50.15	50.23
s_e	EE	0.0472	0.0463	0.0369	0.0373	0.0314	0.0319
s_n	NE	0.151	0.15	0.117	0.145	0.111	0.108
δ^{eu}_{ss}	EU	0.0154	0.0153	0.0118	0.0122	0.0093	0.0099
δ^{en}_{ss}	EN	0.0384	0.0323	0.0324	0.0337	0.0357	0.0375
δ^{nu}_{ss}	NU	0.0244	0.0228	0.0209	0.0208	0.0155	0.0149
δ^{un}_{ss}	UN	0.229	0.213	0.253	0.245	0.258	0.264

Notes: The upper table, "target value," displays the targeted value for each matched moment. Target values (ur_{ss}, EE, NE, EU, EN, NU, UN) are computed as the average of the final 12 months in the series for each episode, using the BLS series for unemployment and employment state gross flows as in Shimer (2012). The lower table contains the parameter values for each episode.

Table 4: Other parameters that vary by recovery

Parameter	Target	1975	1982	1992	2003	2009	2020
a	$\log \mathrm{UE}^3_{ss} - \log \mathrm{UE}^1_{ss}$	4.19	4.35	7.29	4.94	6.65	6.89
b	$b = w_{ss}(0)$	14.1	14.1	14.1	13.1	13.7	13.7

Notes: Parameter values of the meeting efficiency parameter a and the flow value of non-employment b for every recovery episode.

Table 5: Start and end points of recoveries

Recovery	Start point	Unemployment rate	End point	Unemployment rate
2020	2020m4	14.7%	2023m3	3.5%
2009	2009m10	10.0%	2020m2	3.5%
2003	2003m6	6.3%	2007m11	4.7%
1992	1992m6	7.8%	2000m12	3.9%
1982	1982m12	10.8%	1990m7	5.5%
1975	1975m3	9.0%	1979m12	6.0%

Notes: Start and end points of recovery episodes. I define the onset of each recovery as the period of peak unemployment according to the BLS. For the final period, I select a month when unemployment has stabilized.

B Appendix

B.1 Extended discussion of the matching setup

In this appendix, I include an extended discussion of the matching setup in this paper. I begin by highlighting the similarities and differences between the matching mechanism in Butters (1977) and the mechanism described in this paper. As will be clear, Butters and many of the papers adapting similar urn-ball-style models to settings of worker selection, such as Barnichon and Zylberberg (2019) or Bradley (2022), can be understood as a "many-to-one" version of the many-to-many matching setup outlined in this paper. Motivated by this insight, I continue with a discussion of the advantages of modeling two-sided rather than one-sided multiplicity of encounters when analyzing selection mechanisms at business cycle frequencies. Lastly, I discuss the differences between my setup and those in Wolthoff (2018) and Birinci, See and Wee (2024), two other models that feature two-sided multiplicity of encounters.

B.1.1 Comparison with existing literature on multiple encounters

The idea that workers or firms may face multiple encounters and select among them goes back at least to Butters (1977). In the matching setup of Butters (1977), there is a finite mass of buyers and an indefinite mass of sellers. Buyers post ads at random in buyers' mailboxes, which yields a Poisson distribution of ads in each mailbox. Each ad specifies a price. Each buyer buys one unit of the good from the lowest-price seller that they encounter.

Since Butters assumes the cost of posting ads to be linear for each seller, the mass of sellers is irrelevant for the outcomes of interest. The analogue of this assumption in a labor market setting would be a free entry condition that specifies a linear cost of posting vacancies for an indeterminate mass of potential entrants. Thus, an "ad" is the analogue for a "vacancy" and Butters's model can be understood as a many-to-one matching setup with wage posting in which workers encounter multiple vacancies and vacancies encounter at most one worker.

The matching model developed in this paper makes two important changes to this setup. First, since the goal of this paper is to analyze selection of workers, each vacancy (rather than each worker) has to be given the ability to choose among multiple encounters. This requires reversing the seller and buyer labels in Butters and assuming free entry on the firm side. I further simplify the model by assuming that wages are bargained rather than posted. These changes alone yield a "many-to-one" matching model very similar to the one studied in Barnichon and Zylberberg (2019), where each job seeker sends a single application to vacancies and each vacancy selects from multiple candidates.

Second, I allow for multiple encounters on *both* sides of the market, workers and firms. This feature is shared with the models in Wolthoff (2018) and Birinci, See and Wee (2024). However, unlike both Wolthoff (2018) and Birinci, See and Wee (2024), in my model firms rank workers identically and worker differences are fixed rather than idiosyncratic. This is important in order to think through the selection issues considered in this paper. I argue below that this is a useful generalization of this class of model that may also prove useful for other settings.

B.1.2 Two-sided versus one-sided multiplicity of encounters

The model's matching stage allows for two-sided multiplicity of encounters. An alternative assumption would be to allow for only one match per worker similar to Barnichon and Zylberberg (2019). I now discuss why a two-sided multiplicity of encounters is desirable when analyzing business cycle fluctuations and selection. The key argument is that the degree of selection is naturally limited with one-sided multiplicity.

To set up the argument, consider an environment with L_t job seekers and V_t vacancies, which are connected through M_t encounters, and focus on the setting with one-sided multiplicity where each worker encounters at most one vacancy.

First, note that the number of successful matches $\tilde{M}_t$ is bounded above by V_t. For this reason, the average job finding rate $\tilde{M}_t/L_t$ is bounded above by V_t/L_t, which must be large enough to accommodate realistic job finding rates.

However, with one-sided multiplicity, any job seeker can have at most one encounter, and thus the number of encounters per vacancy M_t/V_t is bounded above by L_t/V_t, the inverse of the upper bound for the job finding rate. Importantly, the number of encounters per vacancy governs the degree of competition faced by workers and therefore the relationship between worker rank and the level and cyclicality of their individual job finding probabilities. Thus, these inequalities show that there is an upper bound to how selective vacancies can be given realistic job finding rates. The same is not true for an environment with two-sided multiplicity.

Note that to derive these bounds, we have used weak inequalities that generally will not hold with equality in a given application. How close one gets to these bounds depends on assumptions such as the functional form of M_t as a function of V_t and L_t. In a given application, these constraints may thus easily become more restrictive than they appear to be here. For this reason, two-sided multiplicity as outlined in this paper is a preferable modeling approach and a useful generalization in a business cycle setting.

B.1.3 Fixed versus idiosyncratic differences across workers

Two other models of two-sided multiple encounters are Wolthoff (2018) and Birinci, See and Wee (2024). The main difference between these models and the model outlined in this paper is that both of these models consider a setting in which match quality is given by a purely idiosyncratic draw and there are no permanent differences across workers. In my model on the other hand, worker heterogeneity is permanent rather than idiosyncratic. This is the key feature that gives rise to the selection mechanism in which some types of workers are *repeatedly* outranked by others and composition effects affect vacancy posting incentives. This type of selection requires permanent differences across different types of workers, which is a feature not nested within the aforementioned frameworks.

B.1.4 Other applications for the matching framework

The model may prove useful for adaptation in future work. One possible application of this type of matching model could be assortative matching: The many-to-many matching setup is in principle very adaptable to other settings where workers *and* firms are heterogeneous. With fixed firm heterogeneity, the matching stage generates natural assortative matching with the degree of assortativeness pinned down by the meeting efficiency parameter a. Thus, this generalization may prove useful in future work focusing on such issues as the impact of changes in matching technology on inequality and assortative matching.

Another strand of work, originating in Merz (1995) and Andolfatto (1996), has focused on the ability of search models embedded in RBC models to generate the empirical co-movement of various variables. The matching model in this paper could in principle be embedded in a model of this type and provide another horse in the race when it comes to matching empirical co-variances of aggregate variables and the persistence of shocks. The model feature that $\{\lambda_t\}_{t=0}^{\infty}$, a vector of scalars, determines the entire cross-section of job finding rates over the cycle, makes this application much easier to handle from a technical perspective. For example, even though the model with many-to-many matching does not allow for closed-form solutions thanks to the differential equation (8), it could nonetheless be solved by computing impulse responses as in this paper and then using standard linearization techniques to compute the responses of endogenous variables to a sequence of exogenous aggregate shocks. In terms of calibration, the present paper gives some guidance on how to choose cross-sectional parameters, but cross-sectional data on wages and differential job finding rate is needed to discipline cross-sectional parameters.

B.2 Random versus directed search

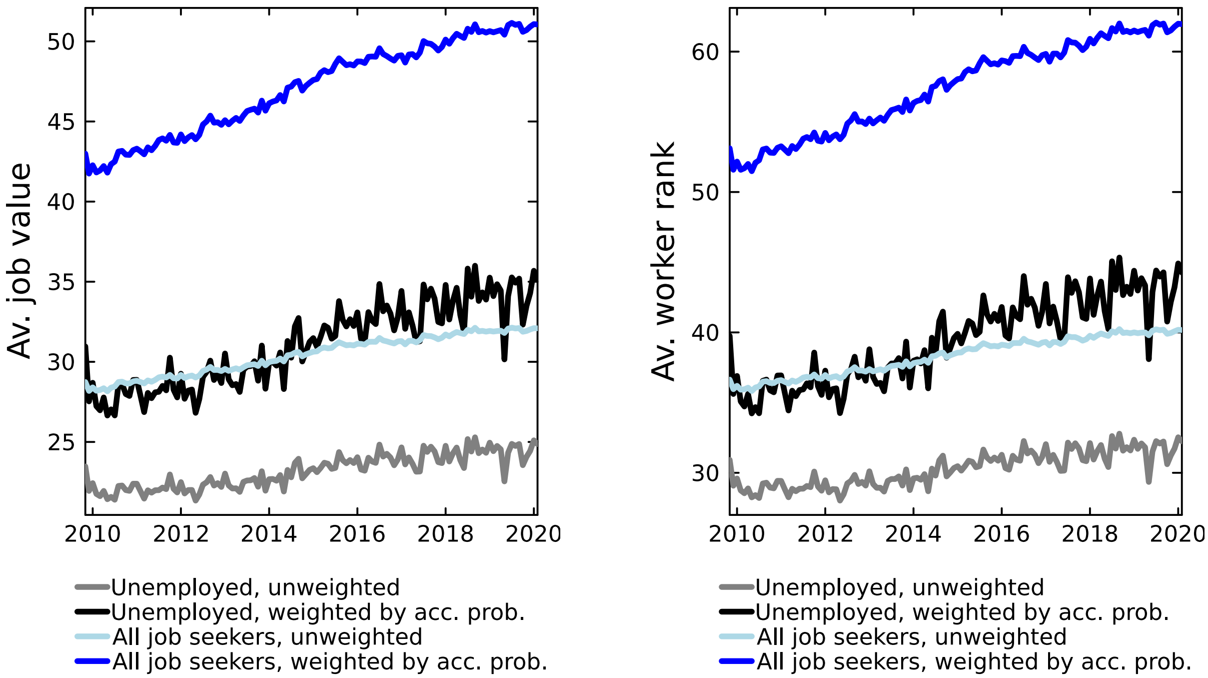
The matching mechanism relies on the assumption that encounters are randomly allocated, meaning that workers have no control over the firms that they encounter. This form of random search stands in contrast to the literature on directed search, which assumes that workers have perfect control over the target of their application.

In a traditional directed search model with wage posting and observable types, workers of different rank would separate their search into different markets, with queue lengths and wages equalizing to make firms indifferent between posting in different markets. An example of this kind of setup is Menzio and Shi (2011). In a model like this, it is still possible to generate differential cyclicality of UE rates in response to productivity shocks - for example, by making a productivity shock diminish the difference between the average match productivity and the value of unemployment more for low-type workers than for high type workers. Gregory, Menzio and Wiczer (2024) study a mechanism that works partly through this channel. However, the type of friction that gives rise to selection in our model is absent from this framework.

It is possible to imagine variants of the model presented in this paper that feature some type of direction of applications. Some of them are isomorphic to the model presented here. In particular, when workers are assumed to be truly indifferent across vacancies (i.e. high rank vacancies are interpreted as "being the first to pick up the phone") and this ranking is revealed only after application decisions are made, then the fact that workers can direct their search does not matter, and there is a symmetric non-coordination equilibrium in which a given worker uniformly randomizes over applying to each given vacancy. This is no longer true in a case in which the vacancy ranking is a true preference or productivity ranking and known ex-ante. Such a model would be similar to the assignment model in Oh (2024), with positive assortative matching as the outcome. However, it is unclear whether the differential level and cyclicality in job finding probability would survive in such a setup.

One could also try to introduce a notion of "semi-directed" search into the model by interpreting encounters as informational constraints ("knowing about" a position), and then allowing workers to pick a subset of encounters that they would like to be considered for (the latter being an “application”). In a world without firm heterogeneity, this model would again be isomorphic to the baseline, as workers would be indifferent between jobs. With firm heterogeneity, high type workers would optimally apply to higher type positions, which would increase the assortativeness of matching. Again, the steady state and cyclical properties of relative job finding probabilities are not obvious from this intuition. Solving these types of modifications of the model, which involve their own technical challenges, is a task left to future work.

C Additional figures



(a) Av. job value $J_t(i)$, 2009 recovery

(b) Av. worker rank i, 2009 recovery

Figure 20: Various measures of av. job seeker quality

Notes: Different measures of composition effects predicted by the model during the recovery from the Great Recession as defined in equation (19). Left panel: $\text{Quality}(i) = J_{ss}(i)$. Right panel: $\text{Quality}(i) = i$. “Unweighted” means $\text{SelectionWeight}(i) = 1$; “weighted by acc. prob.” means $\text{SelectionWeight}(i) = \sigma_{ss}(p_L^{ss}(i))/(\int_0^1 \sigma_{ss}(\tilde{p}_L)\, d\tilde{p}_L)$. The label “unemployed” refers to measures with $\text{Prevalence}_t(i) = U_t^-(i)/(\int U_t^-(\iota)\, d\iota)$; the label “all job seekers” refers to measures with

$\text{Prevalence}_t(i) = (U_t^-(i) + s_n N_t^-(i) + s_e E_t^-(i))/(\int U_t^-(\iota) + s_n N_t^-(\iota) + s_e E_t^-(\iota)\, d\iota)$.

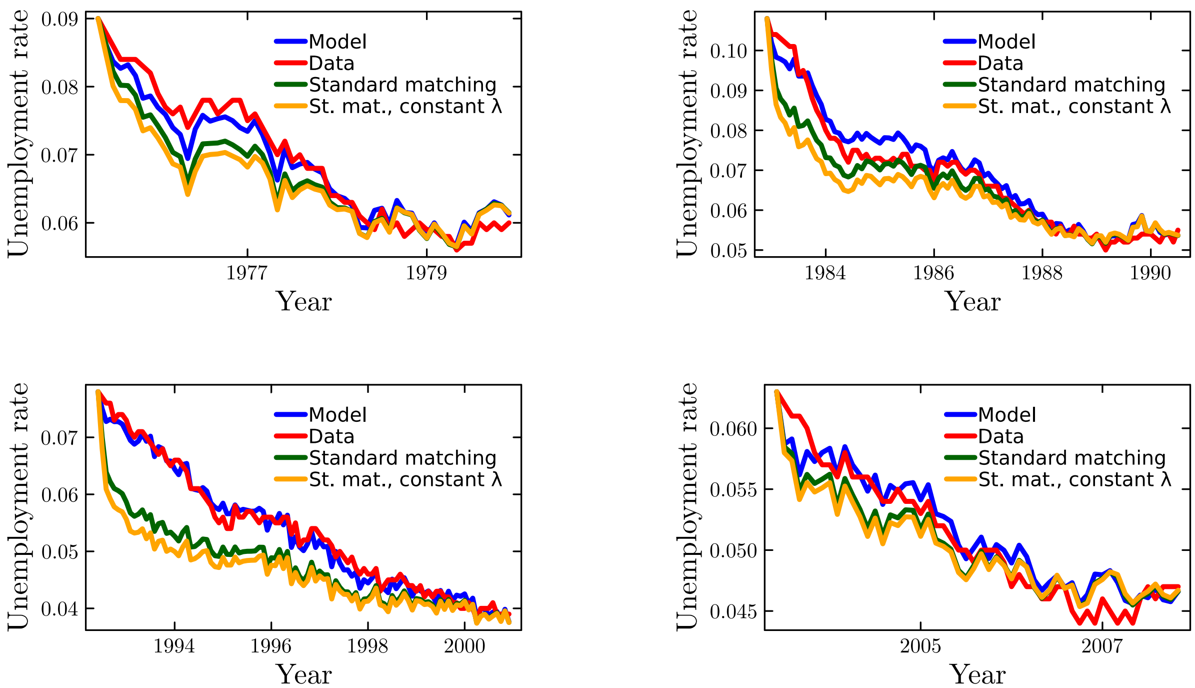


Figure 21: True and simulated unemployment series for other recoveries

Notes: See Figure 15.

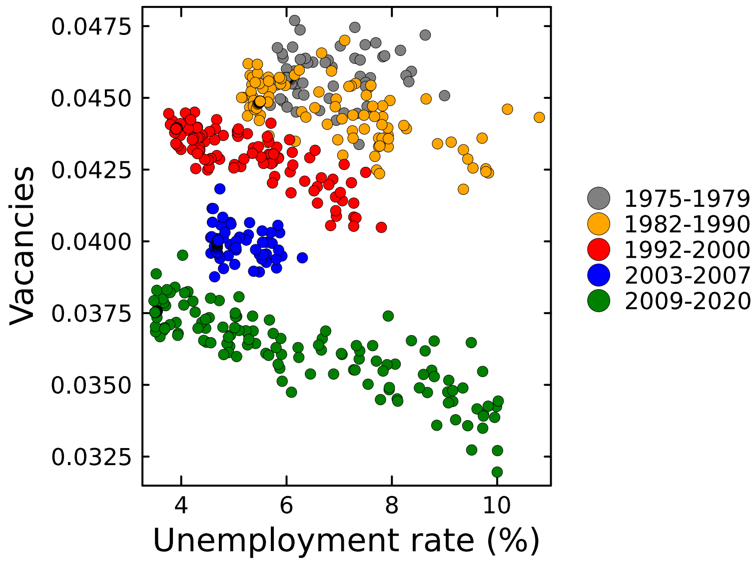
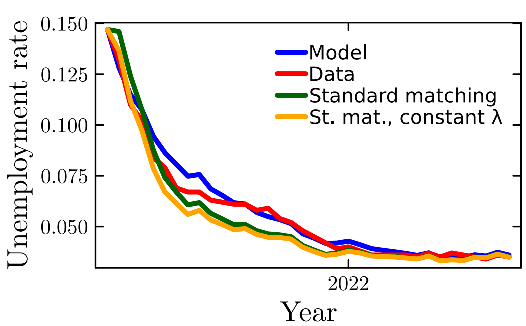
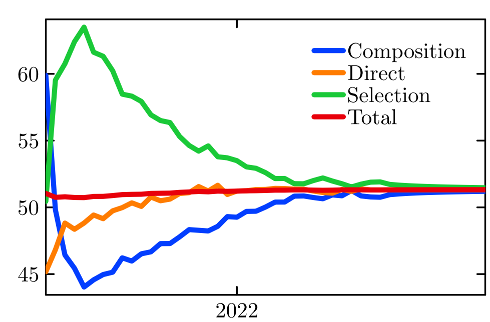


Figure 22: Beveridge curves for modeled pre-COVID recoveries

Notes: Unemployment rate and vacancy stock as predicted by the baseline model across all five pre-COVID recovery episodes.



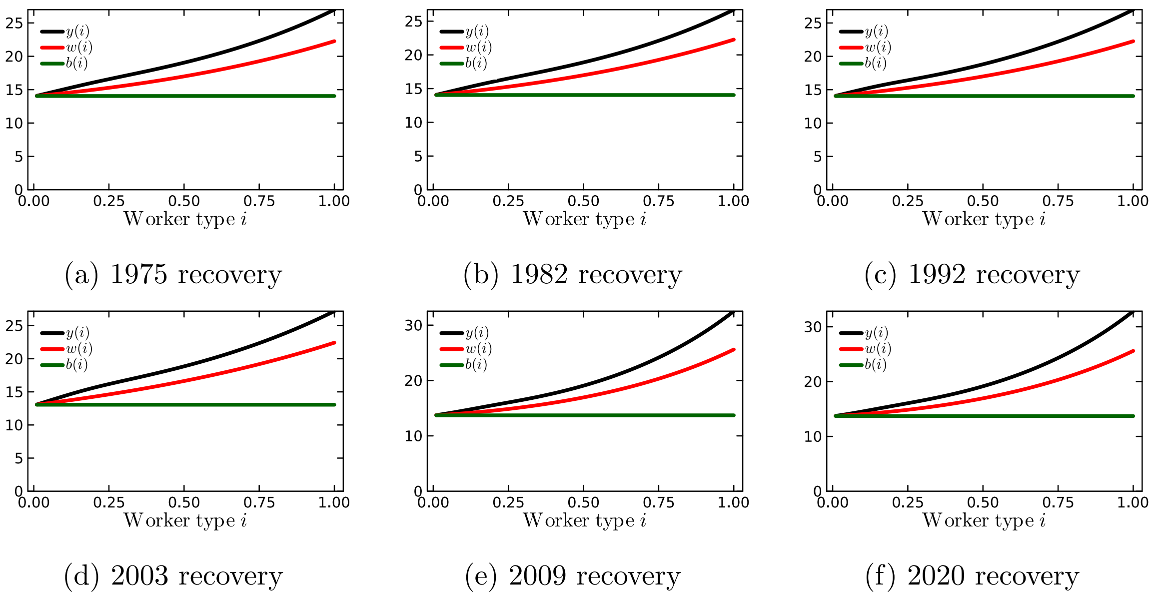
(a) 2020 recovery



(b) Decomposition of J_t

Figure 23: Model predictions for the COVID recovery

Notes: See Figures 15 and 19.



(a) 1975 recovery (b) 1982 recovery (c) 1992 recovery

(d) 2003 recovery (e) 2009 recovery (f) 2020 recovery

Figure 24: Calibrated functions $y(i)$, $w(i)$ and $b(i)$

Notes: Productivity, steady state wages and the flow value of non-employment as a function of rank in the model, by recovery episode. Here, $w_{ss}(i)$ is estimated by projecting the average wage by LJR centile q_{100} onto a second-order polynomial as described in the main text, $b(i) = b$ is assumed to be equal to the lowest steady state wage in the economy for each episode, and $y(i)$ is then inferred for each steady state using the model.

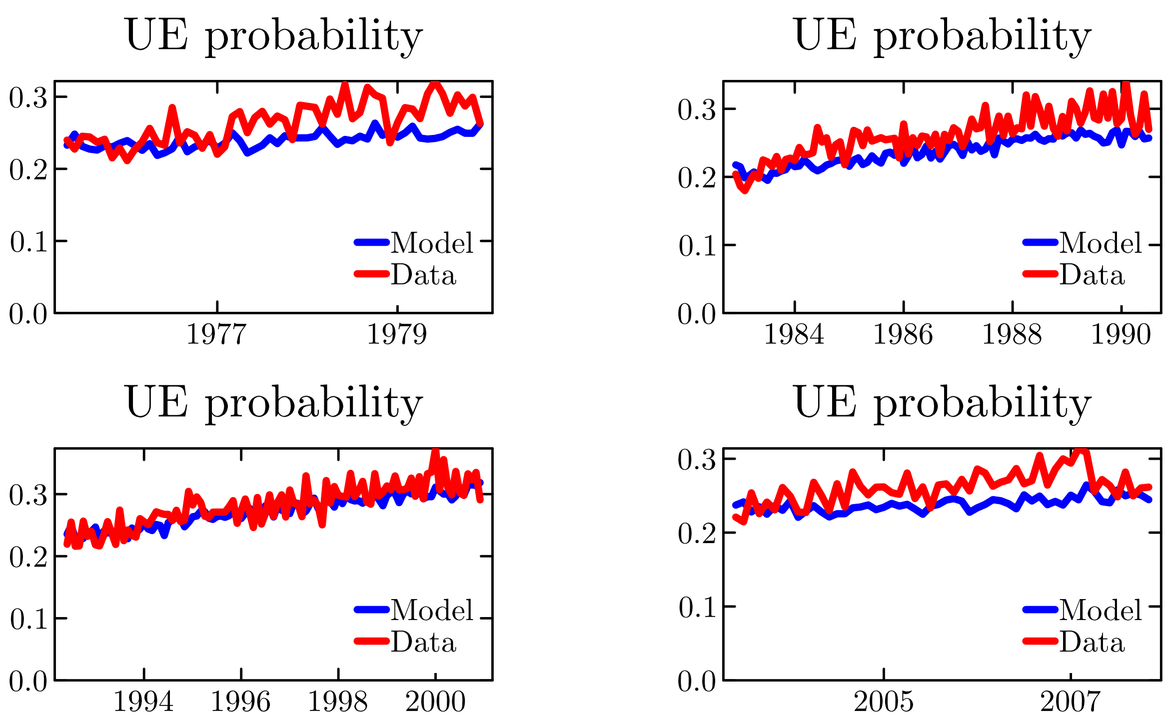


Figure 25: UE transition probability for 1975-2003 recoveries

Notes: See Figure 16.

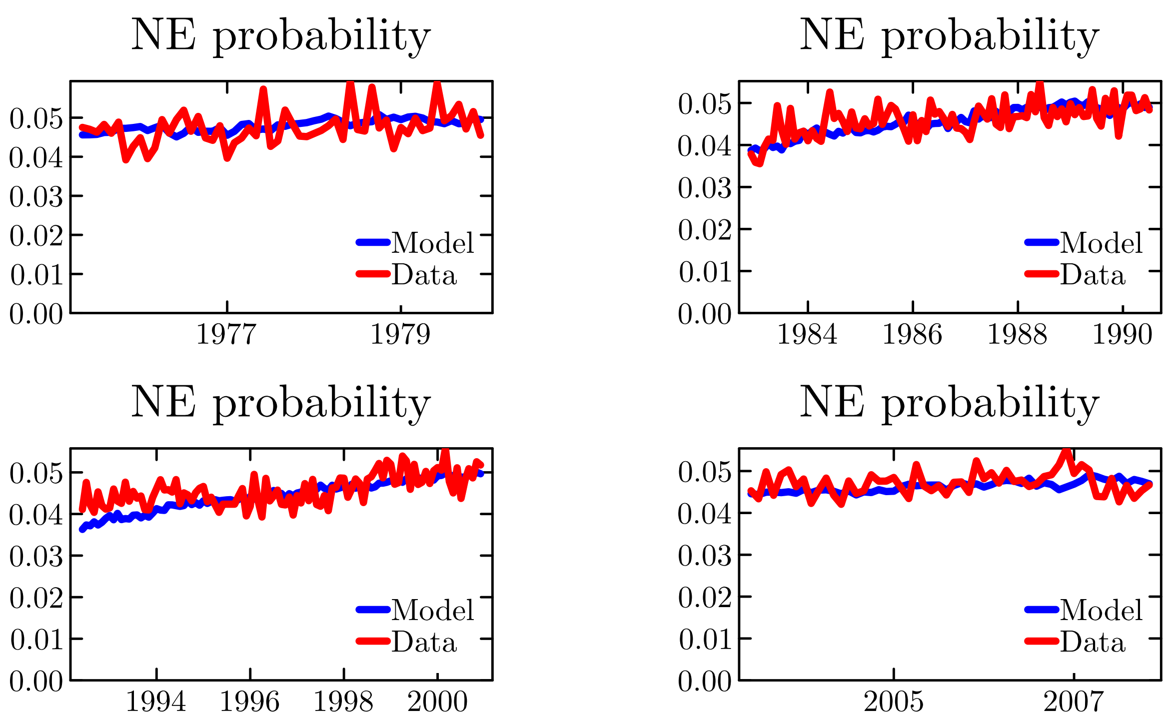


Figure 26: NE transition probability for 1975-2003 recoveries

Notes: See Figure 16.

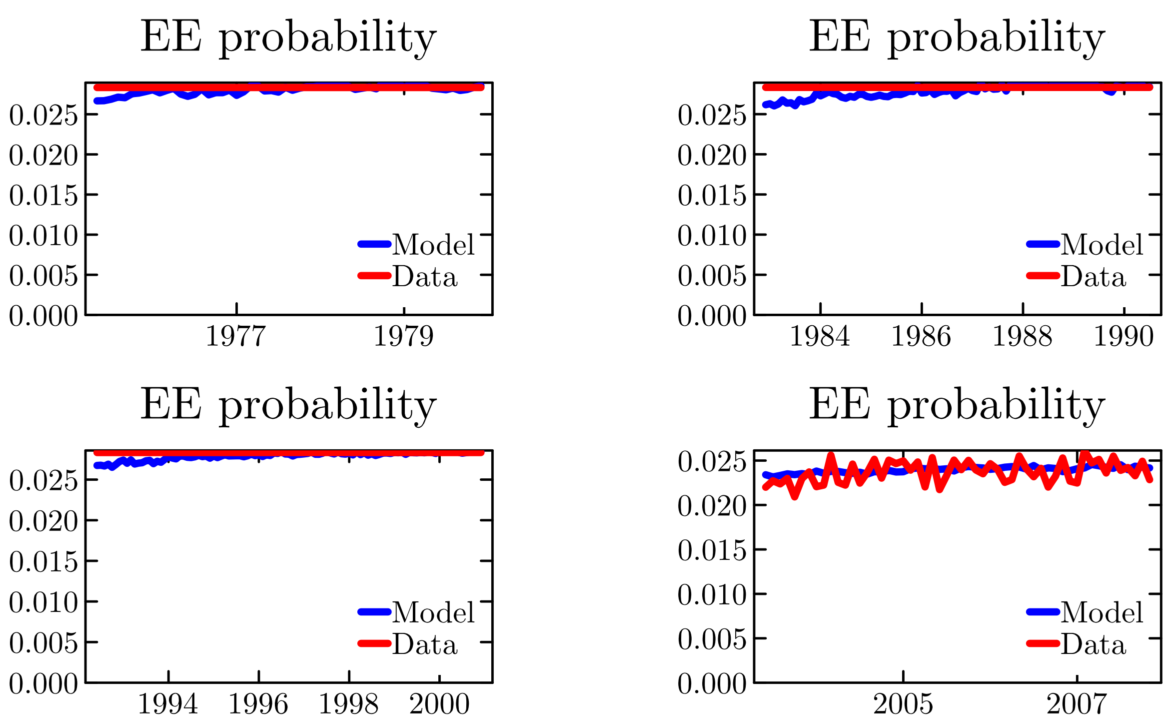


Figure 27: EE transition probability for 1975-2003 recoveries

Notes: The red line for the 1974, 1982, and 1992 recoveries shows the assumed steady state value as real-time data is missing. See Figure 16 for further details.

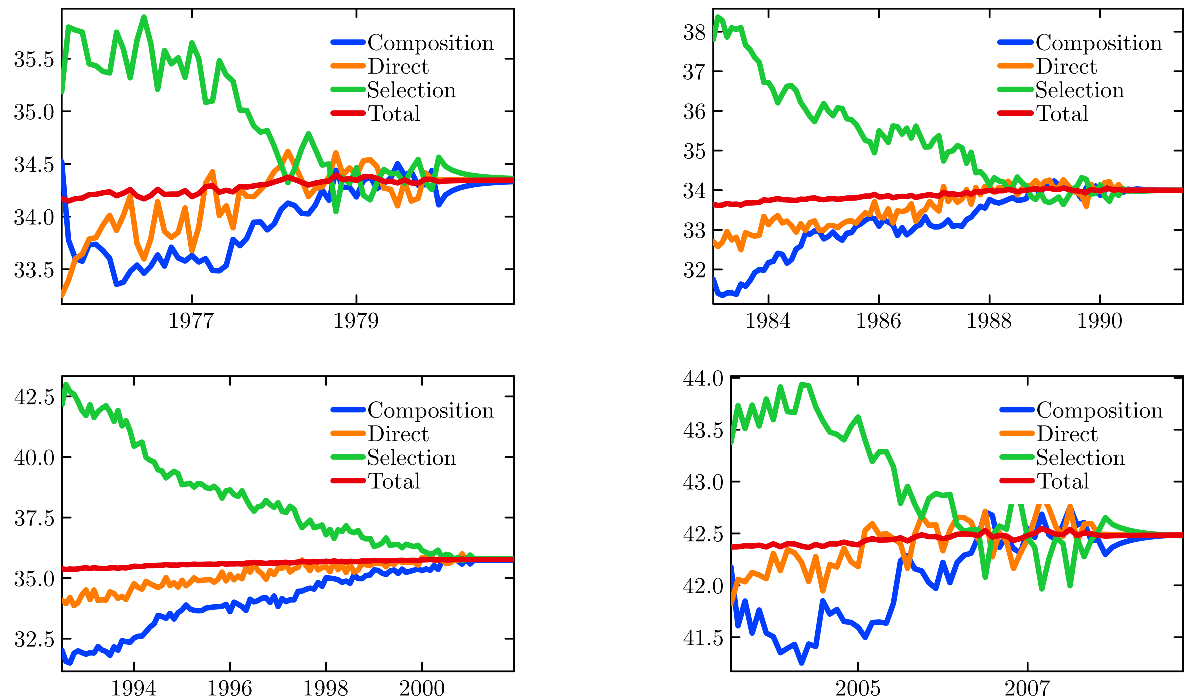


Figure 28: Decomposition of J_t for 1975-2003 recoveries

Notes: See Figure 19.

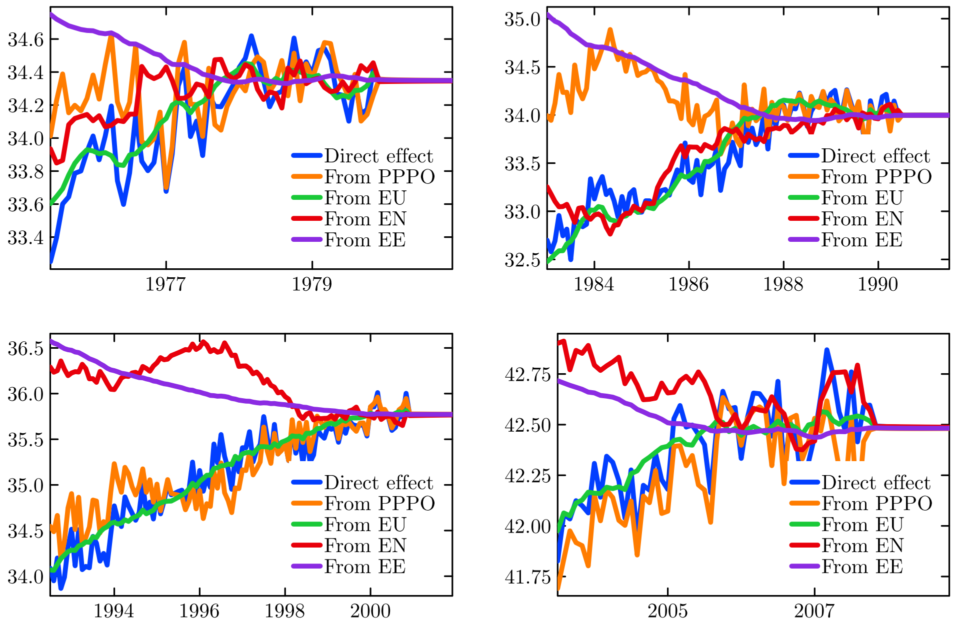
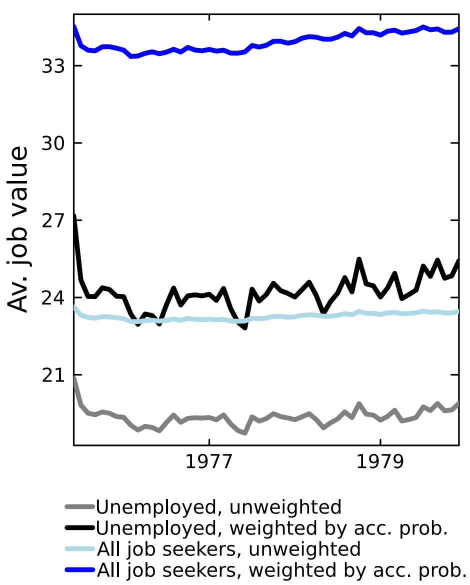
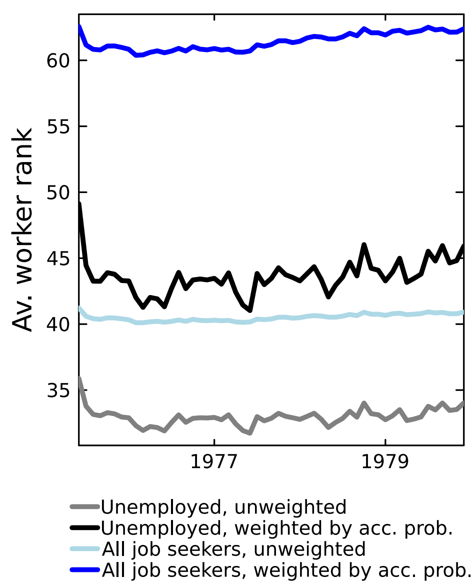


Figure 29: Decomposition direct effect of J_t for 1975-2003 recoveries

Notes: See Figure 19.



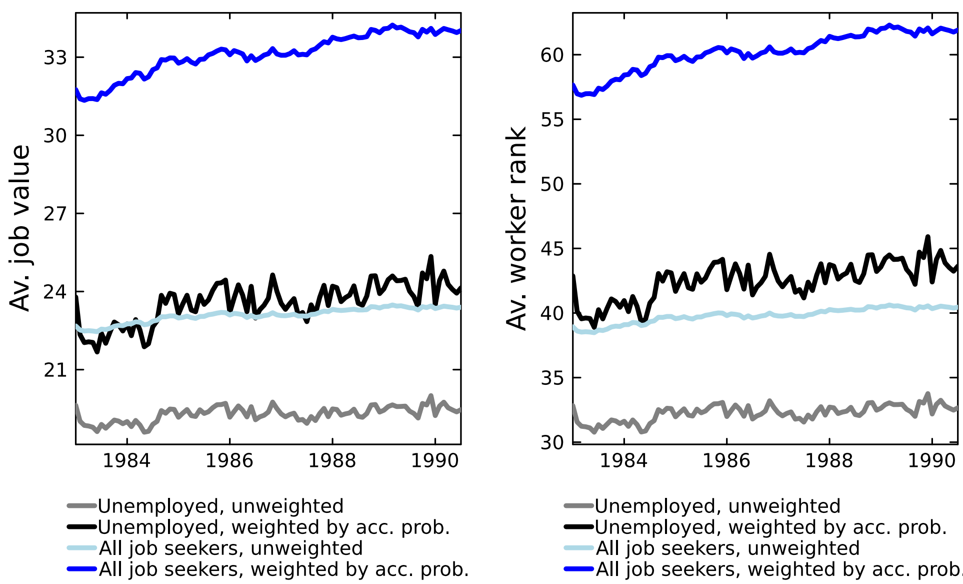
(a) Av. job value $J_t(i)$, 1975



(b) Av. worker rank i, 1975

Figure 30: Various measures of av. job seeker quality, 1975 recovery

Notes: See Figure 20.

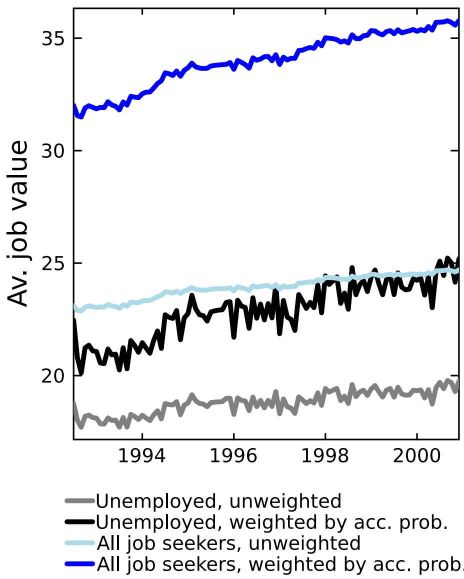


(a) Av. job value $J_t(i)$, 1982

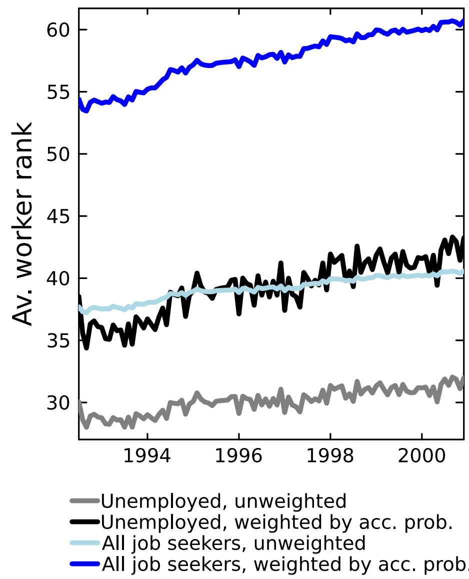
(b) Av. worker rank i, 1982

Figure 31: Various measures of av. job seeker quality, 1982 recovery

Notes: See Figure 20.



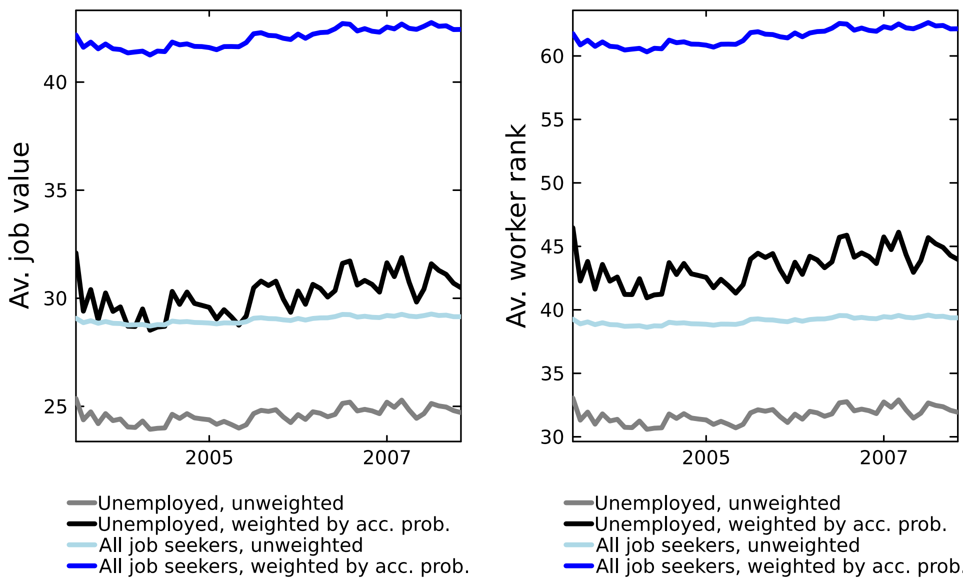
(a) Av. job value $J_t(i)$, 1992



(b) Av. worker rank i, 1992

Figure 32: Various measures of av. job seeker quality, 1992 recovery

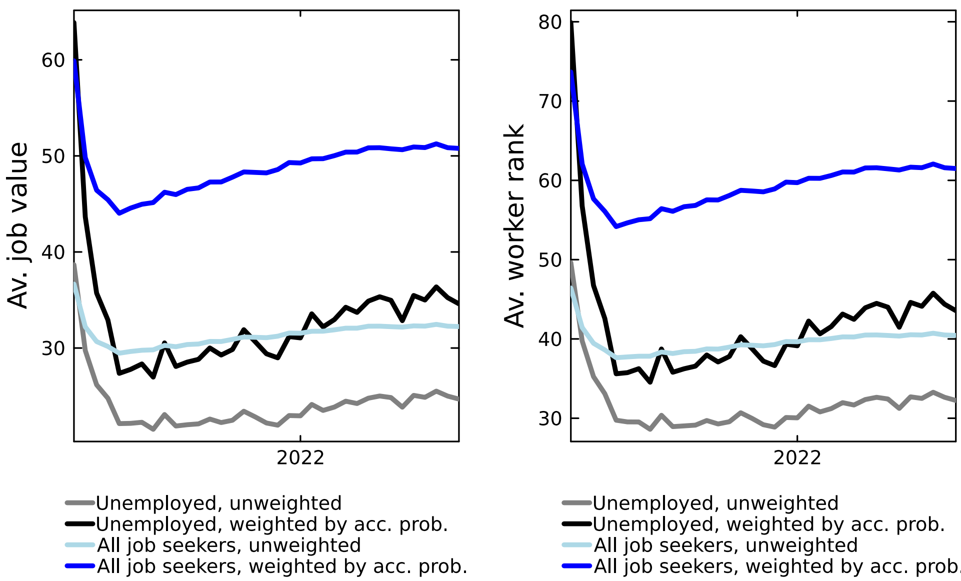
Notes: See Figure 20.



(a) Av. job value $J_t(i)$, 2003 (b) Av. worker rank i, 2003

Figure 33: Various measures of av. job seeker quality, 2003 recovery

Notes: See Figure 20.



(a) Av. job value $J_t(i)$, 2020 (b) Av. worker rank i, 2020

Figure 34: Various measures of av. job seeker quality, 2020 recovery

Notes: See Figure 20.