The Alpha Beta Gamma of the Labor Market*

Victoria Gregory
FRB Saint Louis

Guido Menzio
New York University and NBER

David Wiczer
Stony Brook University and NBER

October 2021

Abstract

Using a large panel dataset of US workers, we calibrate a version of the search-theoretic model of the labor market of Menzio and Shi (2011) where workers are heterogeneous with respect to the parameters governing their employment transitions. The model is calibrated with a version of the 2-stage Grouped-Fixed Effect of Bonhomme, Lamadon and Manresa (2021), where heterogeneity is approximated by discrete latent types and type-specific parameters are calibrated by matching type-specific moments. The model is used to measure the effect of aggregate productivity shocks. As the unemployment of the marginal type is sensitive to the shock and adjusts slowly, the response of aggregate unemployment is large and persistent.

JEL Codes: E24, O40, R11.

Keywords: Unemployment Fluctuations, Grouped Fixed Effects, Block Recursivity.

*We are grateful to Steve Davis, Jean Flemming, Jim Hamilton, Loukas Karabarbounis, Simon Mongey, Iacopo Morchio, Giuseppe Moscarini, Andi Mueller, Antonella Trigari, and audience members at the NBER Summer Institute, the Minnesota Macro Workshop, SITE, and seminars at UT Austin, UW Madison, NYU and Barcelona GSE for many useful comments. We thank Bob Hall and Mariana Kudlyak, with whom we had numerous conversations about our related projects. We are especially grateful to Fabian Lange, Thibaut Lamadon and Elena Manresa for their help in shaping the analysis of workers' heterogeneity. The views expressed here are solely those of the authors and do not necessarily reflect those of the Federal Reserve Bank of St. Louis or the Federal Reserve System. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation.

1 Introduction

The standard tool for studying cyclical unemployment fluctuations is the equilibrium search-theoretic model of the labor market developed by Diamond, Mortensen and Pissarides (e.g., Pissarides 1985 or Mortensen and Pissarides 1994). In nearly all of its formulations and applications, the model assumes that workers are ex-ante identical with respect to their pattern of employment transitions—in the sense that their transitions are all realizations of a common stochastic process. Yet, empirical evidence—even casual one—suggests that workers move across employment states according to heterogeneous transition processes. Indeed, the rate at which workers move between employment and unemployment is known to be different for workers with different education. There is evidence that the majority of unemployment spells originates from the same minority of workers (e.g., Morchio 2019). And there is evidence of unobserved heterogeneity in the rate at which workers move from unemployment to employment (e.g., Mueller, Spinnewjin and Topa 2019). Naturally, though, evidence of heterogeneity in the transition process of workers does not automatically imply that the "representative worker" assumption is inappropriate for studying aggregate unemployment fluctuations and other macro-labor phenomena.

In this paper, we want to use a large panel dataset to provide a systematic assessment of workers' heterogeneity with respect to their pattern of employment transitions and, then, we want to quantify the role played by workers' heterogeneity in shaping the cyclical behavior of the aggregate labor market. Two technical hurdles stand in the way of this project. First, we need to find a parsimonious way to measure worker heterogeneity, since estimating individual-specific parameters through maximum likelihood would be computationally too burdensome on a panel dataset that covers many workers for many years. Second, we need to find a way to solve for the out-of-steady-state dynamics of a model in which workers are heterogeneous not only with respect to their current employment state, but also with respect to the parameters that affect their employment transition process.

In order to recover worker-specific parameters, we follow the two-stage Grouped Fixed Effects (GFE) methodology developed by Bonhomme, Lamadon and Manresa (2021). In the first-step, we discretize workers' heterogeneity using the k-means algorithm. The output of the algorithm is an assignment of individual workers to types and means of type-specific outcomes. In the second-step, we estimate the type-specific parameters of a search-theoretic model of the labor market in the spirit of Menzio and Shi (2011) by matching type-specific moments. As shown in Bonhomme, Lamadon and Manresa (2021), this two-step procedure produces estimates of the type-specific parameters that are consistent—in the sense that they recover the individual-specific parameters in the limit where the number of workers, the number of observations per worker, and the number of types grow large. In order to analyze the implications of workers' heterogeneity for aggregate labor market dynamics, we exploit the fact that the equilibrium of the model is block-recursive—in the sense that the agents' value and policy functions depend on the aggregate state of the economy only through the exogenous shocks and not through the entire distribution of workers across types and employment states.

In the first part of the paper, we document and discretize the extent of workers' heterogeneity with respect to their patterns of employment transitions. We access the Longitudinal Employer-Household Dynamics (LEHD) dataset between 1997 and 2014 and observe the history of employment transitions for over 500,000 individual workers. For each individual worker, we record the time spent in unemployment, the distribution of duration of different unemployment spells and the distribution of duration of different jobs. Using the k-means algorithm, we assign individual workers to types based on the similarity of their pattern of employment transitions. We find that workers' heterogeneity is well-approximated by 3 types of workers, which we label α, β and γ. Workers of type α are the majority of the population. These workers are most likely to move from unemployment to employment within a quarter and, once they become employed, they are most likely to keep their job for more than 2 years. Workers of type γ are a small fraction of the population. These workers are most likely to remain unemployed for more than 1 year and, once they become employed, they are most likely to leave their job within 1 year. The pattern of employment transitions for β-workers is between the pattern for α and for γ-workers. Interestingly, we find that a worker's type is essentially orthogonal to the worker's demographic characteristics and industry—a finding that suggests that grouping workers by demographics and industry is not a good way to approximate heterogeneity in transition patterns.

In the second part of the paper, we develop an equilibrium model of workers' transitions across employment states. The model is similar to Menzio and Shi (2011) and Menzio, Telyukova and Visschers (2016). Firms spend resources to open vacancies and they advertise the terms of trade offered to workers hired to fill them. Workers spend time seeking vacancies and direct their search towards jobs offering particular terms of trade. In deciding where to search, workers trade off the probability of finding a job with the generosity of the terms of trade offered by the job, and the resolution of the trade-off depends on the worker's employment position. The quality of a particular firm-worker match is random and observed only after the match is consummated. If the quality is low enough, the worker moves into unemployment to find a new job as quickly as possible. If the quality of the match is in an intermediate range, the worker keeps searching while staying on the job. If the quality of the match is high enough, the worker stops searching altogether. Hence, the model generates endogenous transitions between unemployment and employment (UE rate), between employment and unemployment (EU rate) and across employers (EE rate). The transition probabilities differ across types because a worker's type affects his ability to search, his baseline productivity, the distribution from which he samples the quality of his matches, and the speed at which he discovers the quality of his current match.

We calibrate the type-specific parameters of the model by matching type-specific moments, such as the distribution of unemployment spell durations, the distribution of job durations, the unemployment rate, the average earnings, etc... The calibrated model matches the type-specific moments quite well and offers a structural interpretation to the observed pattern of employment transitions for different types of workers. For instance, α-workers move quickly from unemployment to employment because their expected gains from trade are large, and they are likely to keep the same job for a long time because the distribution of match qualities from which they sample has low variance—which implies that they are unlikely to leave a job in order to sample another one. In contrast, γ-workers move slowly from unemployment to employment because their expected gains from trade are small, and they are likely to leave a job within 1 year because the distribution of match qualities from which they sample has a thick right tail—which implies that they keep sampling jobs until they find one at the very top of the quality distribution.

We validate the model by testing its predictions with respect to two classic micro-phenomena. First, we consider the earnings losses of displaced workers—workers who lose a job that they held for more than 3 years. These losses contain information on the search capital accumulated by workers in jobs that survive for multiple years, as well as on the speed at which search capital is rebuilt after a displacement episode. In the data, we find that the earnings losses of displaced workers are large and persistent on average, but much more so for γ-workers than for α-workers. We show that the model reproduces very well the magnitude and persistence of the earnings losses for different types of workers. Second, we consider the relationship between unemployment duration and UE rate. In the data, we find that the UE rate declines sharply with unemployment duration and that the distribution of the unemployment pool tilts towards γ-workers and away from α-workers. We show that the model reproduces well the decline in the UE rate, the composition of the unemployment pool at the beginning of a spell, and its evolution throughout a spell.

In the last part of the paper, we use the calibrated model to measure the effect of aggregate shocks on labor market outcomes. We find that a negative shock to the aggregate component of productivity generates responses in UE, EU and unemployment rates that are very different across different types of workers. The shock leads to a small decline in the UE rate, and to a small and short-lived increase in the EU rate of α-workers. As a result, the increase in the unemployment rate of α-workers is small and transitory. In contrast, the shock leads to a large decline in the UE rate, and to a large and persistent increase in the EU rate of γ-workers. As a result, the increase in the unemployment rate of γ-workers is large and persistent. Intuitively, the UE rate of γ-workers is more sensitive to the shock because the gap between the market productivity and the value of non-market activities is smaller for γ-workers and, hence, the shock reduces their gains from trade by a larger percentage. The EU rate of γ-workers is more sensitive to the shock because γ-workers are more likely to be in marginal matches. The EU rate of γ-workers is affected for a longer time by the shock because displaced γ-workers need to sample several jobs and to go through several EU transitions before finding another stable job.

At the aggregate level, we find that the shock leads to an increase in unemployment that is large and persistent and to a decline in labor productivity that is smaller and more transitory than the shock. In response to a negative shock to the aggregate productivity with a magnitude of 10% and a half-life of 3 years, the increase in the unemployment rate is 7.5 percentage points with a half-life of close to 6 years. Both the magnitude and persistence of aggregate unemployment are largely driven by the behavior of γ-workers, who are a small fraction of the population but control the behavior of aggregate unemployment because they are marginal. In response to the shock, the decline in labor productivity is about 7.5% with a half-life of

2 years. The decline in labor productivity is smaller than the underlying shock because of a double cleansing effect—i.e. the workers who survive the impact of the shock are more likely to be in high-quality matches and are more likely to be high-productivity types. The decline in labor productivity recovers more quickly than the underlying shock because of changes in the composition of the employment pool. Overall, the model implies that unemployment is quite elastic to labor productivity fluctuations and that unemployment fluctuations are much more persistent than labor productivity fluctuations.

We assess the importance of workers' heterogeneity for macro dynamics by measuring the response to the shock in a version of the model with a representative worker. We find that workers' heterogeneity almost doubles the size of the responses of UE, EU and unemployment rates, and that is nearly doubles the half-life of the response of EU and unemployment rate. Intuitively, a representative worker model that is calibrated to match the population-wide pattern of workers' transitions between employment states fails to capture the properties of the search process of γ-workers, which is the process that essentially controls the response of the labor market to aggregate shocks.

We measure the unemployment rate of different types of workers during and after the Great Recession. We find that the unemployment rate increased by only 3 percentage points for α-workers, by 8 percentage points for β-workers, and by a striking 20 percentage points for γ-workers. Moreover, we find that the increase in the unemployment rate was reabsorbed quickly for α-workers (2013), less so for β-workers (2014), and much more slowly for γ-workers—whose unemployment rate was still 10 percentage points higher than before the recession in 2014. The size and persistence of the increase in type-specific unemployment during the Great Recession are qualitatively similar to the predictions of the model in response to an aggregate productivity shock, but the magnitude of the decline in labor productivity is much smaller and much more transitory than as predicted by the model. Presumably, this is evidence that other types of shocks were behind the Great Recession. We show, however, that type-specific productivity shocks that are perfectly correlated but are larger for γ than for α-workers can go a long way in realigning the quantitative predictions of the model with the actual behavior of unemployment and labor productivity.

In this paper, we make a methodological contribution to the macro-labor literature by calibrating a dynamic equilibrium model in which workers are heterogeneous with respect to the stochastic process that controls their transitions across employment states. To accomplish this task, we use a version of the equilibrium model of Menzio and Shi (2011) which can be solved in the presence of aggregate shocks in spite of workers' heterogeneity, we access a panel dataset covering more than half a million of individuals for more than 15 years, and we use the Grouped Fixed-Effect method of Bonhomme, Lamadon and Manresa (2021) to discretize the observed heterogeneity in workers' transition patters and to estimate the type-specific parameters of the model. Our paper is related to and inspired by recent work by Morchio (2020), Ahn and Hamilton (2020), Hall and Kudlyak (2020) and Karahan, Ozkan and Song (2019), as well as by some earlier work by Pries (2008). Morchio (2020) documents that most unemployment spells originate from the same individuals—thus bringing evidence of heterogeneity in the workers' stochastic processes. Karahan, Ozkan and Song (2019) document systematic differences in the speed at which workers climb the wage ladder. Ahn and Hamilton (2020) document changes in the relationship between UE rate and unemployment duration before and after the Great Recession—thus suggesting cyclical changes in the type-composition of the unemployment pool. Hall and Kudlyak (2020) use the short panel of the CPS to estimate heterogeneous hazards in and out of unemployment by maximum likelihood. In a prescient paper, Pries (2008) shows—as a proof of concept—that workers' heterogeneity may amplify the response of the labor market to shocks. He shows that amplification is strongest if there are workers who have lower productivity and whose EU rate is more sensitive to shocks. As it turns out, these are precisely the features of γ-workers.

The paper makes a substantive contribution to the macro-labor literature by showing that workers' heterogeneity does matter for labor market fluctuations. From the perspective of a basic search-theoretic model with a representative worker (e.g., Pissarides 1985), the cyclical behavior of the labor market is puzzling. First, unemployment fluctuations are large compared to productivity fluctuations (e.g., Shimer 2005). The literature identified several ways to increase the elasticity of unemployment to productivity shocks: sticky wages (e.g., Hall 2005, Gertler and Trigari 2009, Kennan 2010, Menzio and Moen 2010); small gains from trade (e.g., Hagedorn and Manovskii 2008, Sargent and Lijndquivst 2017); heterogeneous match quality (e.g., Menzio and Shi 2011). In our paper, amplification is generated by γ-workers—who have small expected gains from trade and face great uncertainty with respect to the quality of their matches. Second, unemployment recoveries are slow (e.g., Pries 2004, Bachmann 2007). Mechanisms that are known to increase the persistence of unemployment include: heterogeneous match quality (e.g., Pries 2004); adjustment costs in the stock of vacancies (e.g., Fujita and Ramey 2007); and a decline in the firms' ability to recall previous employees (e.g., Fujita and Moscarini 2010). In our paper, unemployment propagation is generated by γ-workers—whose search process is slow compared to the average worker. Third, the correlation of unemployment and productivity has been low over the last 35 years (e.g., Gali and Van Rens 2017). In order to address the low correlation between unemployment and productivity, the literature considered non-technological shocks, such as shocks to the discount factor (e.g., Hall 2017, Kehoe, Midrigan and Pastorino 2019, Martellini, Menzio and Visschers 2021), self-fulfilling shocks to expectations (e.g., Kaplan and Menzio 2016), and correlated equilibria (e.g., Golosov and Menzio 2020). In our paper, a weak correlation between unemployment and productivity can be generated by type-specific shocks.

2 Documenting and Discretizing Heterogeneity

The aim of the paper is to estimate a search-theoretic model of the labor market in which workers differ with respect to fundamental parameters that shape their pattern of transitions between employment, unemployment and across employers. To this aim, we follow a version of the two-step Grouped Fixed Effects (GFE) estimation method of Bonhomme, Lamadon and Manresa (2021). In the first step, we group workers into a discrete number of types based on their similarity with respect to their pattern of transitions across employment states. In the second step, we introduce a theoretical model of workers' transitions across employment states based on Menzio and Shi (2011) and calibrate its type-specific parameters by matching type-specific empirical moments. As established in Bonhomme, Lamadon and Manresa (2021), the two-step GFE method provides consistent estimates of individual-specific parameters.[1] In this section, we carry out the first step. In Section 2.1, we describe the administrative data that we use to document heterogeneity in workers' employment transitions. In Section 2.2, we describe the algorithm used to group workers into types. In Section 2.3, we describe the defining features of different types.

2.1 Data

Our empirical analysis is based on data from the Longitudinal Employer-Household Dynamics (LEHD) program. The LEHD contains information about individual employment histories, including an identifier for the individual, an identifier for the individual's employer (a state-level SEIN), and the quarterly earnings of the individual from each employer (as measured by pre-tax labor earnings). The LEHD does not report employment in the military or in the federal government, self-employment, contracting work, or any other form of employment that is not covered by Unemployment Insurance.[2]

We have access to the employment histories between 1997 and 2014 for a 2% random sample of individuals from 17 States, including California, Illinois and Texas. Since we are interested in individuals with a strong attachment to the labor force, we purge our sample from all workers who have an earning gap of more than 2 years between two consecutive employment episodes. The purged sample contains about 692,000 unique individuals, or about 0.5% of the US labor force and 0.65% of the US private sector employment. In light of our inclusion restriction, we will refer to non-employment spells as unemployment spells, and we will refer to the fraction of non-employed individuals as the unemployment rate. Obviously, though, our definition of unemployment does not coincide with the official definition of unemployment by the Bureau of Labor Statistics (BLS).

In order to deal with individuals who are entering and exiting the labor force and with censored spells, we create a 2-year window at the beginning of the reference period 1997-2014 (i.e. 1997-1998) and at the end of the reference period (i.e. 2013-2014). If an individual was employed in the first quarter of 1999, we know whether his job lasted more than 2 years (which is the highest bin that we use for classifying job durations) or when it did start (because of the

[1] Hall and Kudlyak (2019) follow a different approach. They maximize the likelihood of each individual history with respect to type-specific parameters and the fraction of workers in each types. Whether one should use maximum likelihood or GFE depends on the available data. If one has access to a small dataset, in which there are very few observations per worker and few workers, maximum likelihood is more appropriate. This is the case in Hall and Kudlyak (2019), who have only 8 observations per individual. On the other hand, if one has access to many observations per worker and many workers, one can appeal to the asymptotic properties of GFE and avoid the computational burden of maximum likelihood. This is our case, as we have up to 56 observations per worker and more than half a million of workers.

[2] See Abowd et al. (2009) for more details about the LEHD.

2-year window at the beginning of the reference period). In either case, we start the record of the individual at the beginning of that job. If an individual was unemployed in the first quarter of 1999, we know whether his unemployment spell lasted more or less than 2 years. If it lasted less than 2 years, we start the record of the individual with the beginning of that unemployment spell. If it lasted more than 2 years, we have no record of prior employment for the individual and we start his record from his first job in the reference period. That is, we assume that the individual was out of the labor force prior to his first job.

Symmetrically, if an individual was employed in the last quarter of 2012, we know whether his job lasted more than 2 years (since we track the worker until the end of 2014) or when the job ended. In either case, we end the record of the individual with the end of that job. If an individual was unemployed in the last quarter of 2012, we know whether his unemployment spell lasted more than 2 years—in which case we stop the record of the individual with the end of the last job in the reference period—or less than 2 years—in which case we stop the record of the individual at the end of the unemployment spell.

Having determined the start and end date of the employment record of each individual, we measure the duration of each of his jobs and each of his unemployment spells. We measure the duration of a job as the number of quarters during which the individual reports earnings from a particular employer.[3] We measure the duration of an unemployment spell as the number of quarters during which the individual does not report any earnings. If the individual transits from one employer to another, he may experience a spell of unemployment lasting less than a full quarter. We assume that the worker experiences a short unemployment spell in one of two cases: (a) the individual has only earnings from the first employer in one quarter and only earnings from the second employer in the next quarter; (b) the individual has earnings from both employers, but the total of these earnings is less than the minimum between his earnings in the previous quarter (when the individual is with the first employer) and in the next quarter (when the individual is with the second employer). In each one of these two cases, we impute an unemployment spell of half a quarter. Otherwise, we assume that the worker transited directly from the first to the second employer.

We then classify all the jobs of the individual into one of four bins: jobs lasting no more than 1 quarter; jobs lasting more than 1 and no more than 4 quarters; jobs lasting more than 4 but no more than 8 quarters; and jobs lasting more than 8 quarters. Similarly, we classify the unemployment spells into one of three bins: spells lasting no more than 1 quarter (which includes imputed unemployment spells); spells lasting more than 1 but no more than 4 quarters; and spells lasting more than 4 but less than 8 quarters.

Lastly, we summarize the pattern of employment transitions of an individual by constructing the following statistics: (i) the fraction of jobs lasting less than 1 quarter, between 1 and 4 quarters, between 4 and 8 quarters, and more than 8 quarters; (ii) the fraction of unemployment spells lasting less than 1 quarter, between 1 and 4 quarters, and between 4 and 8 quarters;[4] (iii)

[3]We restrict attention to the individual's primary employer.

[4]If the individual is employed over the entire period, we set the fraction of unemployment spells of any duration at zero.

the total quarters of unemployment as a fraction of the total number of quarters on record; (iv) the total number of different jobs as a fraction of the total number of quarters on record. These statistics paint a picture of the pattern of employment transitions of a particular individual. Statistic (iii) tells us how much time the individual spends in unemployment. Statistic (ii) tells us the distribution of unemployment durations for an individual. Statistic (i) tells us the distribution of job durations for an individual. Statistic (iv) together with (iii) tells us about direct job-to-job transitions.

2.2 Discretizing heterogeneity

We discretize the workers' heterogeneity with respect to their pattern of employment transitions using the k-means algorithm—a standard tool in the machine learning literature (see, e.g., Friedman, Hastie and Tibshirani 2017) that is becoming commonplace also in economics (see, e.g., Bonhomme, Lamadon, and Manresa 2019, 2021). The k-means algorithm discretizes heterogeneity by grouping workers into types based on their similarity.[5] The number of types is chosen using the cross-validation method by Wang (2010).

Let i denote an individual in our sample. Let $s_{1,i}$, $s_{2,i}$, $s_{3,i}$ and $s_{4,i}$ denote the distribution of job durations for individual i. Let $s_{5,i}$, $s_{6,i}$ and $s_{7,i}$ denote the distribution of unemployment durations for individual i. Let $s_{8,i}$ denote the fraction of time spent by individual i in unemployment. Let $s_{9,i}$ denote the number of jobs of individual i per unit of time. All the statistics are expressed as ratios with respect to their population-wide standard deviation. The four statistics describing the distribution of job durations have a weight of 1/4 each, the three statistics describing the distribution of unemployment durations have a weight of 1/3 each, and the remaining statistics each have a weight of 1.

For a given the number J of types, the assignment of individuals to types is a mapping $j(i)$ from an individual $i \in \{1, 2, ...N\}$ to a type $j \in \{1, 2, ...J\}$ that solves the following minimization problem

$$\begin{aligned} &\min_{j(i)} \sum_{j=1}^{J} \sum_{i=1}^{N} \sum_{k=1}^{9} \mathbf{1}[j = j(i)](s_{k,i} - s^*_{k,j})^2, \\ &\text{s.t. } s^*_{k,j} = \frac{\sum_{j=1}^{J} \sum_{i=1}^{N} \mathbf{1}[j = j(i)] s_{k,i}}{\sum_{j=1}^{J} \sum_{i=1}^{N} \mathbf{1}[j = j(i)]} \end{aligned} \tag{2.1}$$

In words, an individual i is assigned to a type j so as to minimize the squared distance between the statistics of individual i and the average statistics for all individuals assigned to type j.

[5]We group workers based exclusively on their similarity with respect to their pattern of transitions between employment and unemployment and across jobs. We do so because heterogeneity in employment transitions has the most direct effect on unemployment dynamics—our main object of interest. For example, the relationship between unemployment duration and the probability of finding a job is directly affected by workers' heterogeneity with respect to their UE transition rate, and not by workers' heterogeneity with respect to their wage. Had we included measures of education, race, gender or earnings in the k-means algorithm, we would have likely grouped together workers who have rather different transitions just because they have similar education, race, gender or wages.

We solve the minimization problem in (2.1) using an iterative process. To initialize the iteration, we select one of the dimensions describing individuals. We rank individuals along the selected dimension and divide them into J types of equal size. That is, individual i_1 is assigned to type 1 if he ranks in the lowest $1/J$ percent of the population along the selected dimension. Individual i_2 is assigned to type 2 if he ranks in the second lowest $1/J$ percent of the population along the selected dimension, etc... Having created an initial assignment $j_0(i)$, we compute the average $s^0_{k,j}$ of statistic s_k for all individuals i assigned to type j. In the n-th step of the iteration, we solve (2.1) using $s^{n-1}_{k,j}$ instead of $s^*_{k,j}$ in the objective function. The solution of (2.1) is an updated assignment $j_n(i)$. Using the updated assignment $j_n(i)$, we compute an updated average $s^n_{k,j}$ of statistic s_k for all individuals i assigned to type j. We continue the process until we reach a fixed point. We check the uniqueness of the fixed-point by using different dimensions to construct the initial assignment $j^0(i)$.

We choose the number of types J using the cross-validation approach proposed by Wang (2010). We divide our sample of individuals into three subsamples, S_0, S_1 and S_2. The subsamples S_1 and S_2 are for training, and each of them accounts for 25% of the sample. The subsample S_0 is for validation, and it accounts for the remaining 50% of the sample. For any $J \geq 2$, we solve (2.1) on the training subsample S_1 and obtain the type-specific averages $s^1_{k,j}$. We also solve (2.1) on the training subsample S_2 and obtain the type-specific averages $s^2_{k,j}$. We then solve (2.1) on the validation subsample S_0 using $s^1_{k,j}$ instead of $s^*_{k,j}$ in the objective function. This gives us an assignment $j_1(i)$ of individuals in subsample S_0. We do the same using $s^2_{k,j}$ and obtain a different assignment $j_2(i)$ of individuals in subsample S_0 to types. We choose J so as to minimize the number of individuals in S_0 who are assigned to different clusters based on $s^1_{k,j}$ and $s^2_{k,j}$, i.e.

$$\min_{J\geq 2} \sum_{i=1}^{N_0} 1[j_1(i) \neq j_2(i)]. \tag{2.2}$$

The logic behind the criterion (2.2) is simple. If J is too low relative to the "true" number of types, the average statistics of the J groups constructed using the training sample S_1 and S_2 are likely to be quite different, as multiple actual types are artificially clustered together. Similarly, if J is too large relative to the "true" number of types, the average statistics of the J groups constructed using the two training sample are likely to be quite different, as one type is artificially split into multiple groups. In either case, the same individuals in the validation sample S_0 are likely to be assigned to different groups based on the average statistics constructed using S_1 or S_2.

2.3 Worker types

Table 1 reports the outcomes of the classification process described above. We identify three different types of workers in our sample, which we dub α, β and γ. Workers of type α represent the majority of individuals in our sample (57%), while workers of type β are 26%, and workers of type γ are 17%.

Different types of workers have very different patterns of labor market transitions. Consider

	α-workers	β-workers	γ- workers
Population share	0.57	0.26	0.17
Job duration			
<1Q	0.139	0.201	0.360
1Q-4Q	0.187	0.233	0.321
5Q-8Q	0.236	0.238	0.191
>8Q	0.439	0.329	0.119
Unemployment duration			
<1Q	0.794	0.390	0.553
1Q-4Q	0.157	0.558	0.315
5Q-8Q	0.049	0.052	0.133
Fraction of time unemployed	0.036	0.096	0.292
Earnings	$11,050	$7,376	$5,370

Table 1: Descriptive statistics for each worker type

the distribution of unemployment spell durations for different types. For α-workers, the fraction of unemployment spells lasting less than 1 quarter is 79% and the fraction of spells lasting more than 1 year is only 5%. For β-workers, the fraction of unemployment spells lasting less than 1 quarter is 39% and the fraction of spells lasting more than 1 year is 5%. For γ-workers, the fraction of unemployment spells lasting less than 1 quarter is 55% but the fraction of spells lasting more than 1 year is 13%. That is, α-workers typically have short unemployment spells and very rarely have long ones; β-workers are less likely to have short unemployment spells but they also have very few long ones; γ-workers are much more likely to experience long unemployment spells compared to α's and β's. The average time spent in unemployment is 3.5% for an α-worker, 9.6% for a β-worker, and 29.2% for a γ-worker.

Next, consider the distribution of job durations for different types. For α-workers, the fraction of job spells lasting less than 1 quarter is 14%, and the fraction of job spells lasting more than 2 years is 44%. For β-workers, the fraction of job spells lasting less than 1 quarter is 20%, and the fraction of job spells lasting more than 2 years is 33%. For γ-workers, the fraction of job spells lasting less than 1 quarter is 36%, and the fraction of job spells lasting more than 2 years is 12%. That is, α-workers are 50% more likely to remain in the same job for more than 2 years than β-workers, and 4 times more likely than γ-workers. Conversely, γ-workers are 70% more likely to leave a job within 1 quarter than β-workers, and 3 times more likely than α-workers.

Overall, our classification of workers into types paints a clear picture. When unemployed, α-workers are likely to find a job quickly and, once they find it, they are likely to keep it for more than 2 years. Unemployed β-workers find a job less quickly than α-workers and, once they find it, they are less likely to keep it for more than 2 years. Unemployed γ-workers are likely to remain unemployed for a relatively long period of time. Once they find a job, they are likely to leave it within 1 year and to return into unemployment. Indeed, for γ-workers only about one job in 10 lasts for more than 2 years. We also find that, conditional on employment, different types of workers have very different earnings. About 11 thousand US$ per quarter for

	Observables		Observables + Transitions	
	Logit coeff. β	Logit coeff. γ	Logit coeff. β	Logit coeff. γ
High school graduate	-0.162***	-0.247***	-0.0697**	-0.0948**
	(0.00937)	(0.0112)	(0.0234)	(0.0314)
Some college	-0.220***	-0.365***	-0.0873***	-0.157***
	(0.00918)	(0.0111)	(0.0229)	(0.0310)
College graduate	-0.345***	-0.591***	-0.116***	-0.104**
	(0.00970)	(0.0122)	(0.0246)	(0.0342)
Female	0.0756***	0.0252**	0.0282	-0.0183
	(0.00613)	(0.00787)	(0.0154)	(0.0218)
Non-white	0.160***	0.151***	-0.0673***	-0.289***
	(0.00613)	(0.00778)	(0.0153)	(0.0215)
Birth year	0.00669***	0.0297***	-0.00403***	-0.00622***
	(0.000248)	(0.000325)	(0.000737)	(0.000990)
Constant	-13.1***	-58.7***	7.62***	10.1***
	(0.487)	(0.640)	(1.45)	(1.95)
N	678,000		678,000	
Detailed industry?	Yes		Yes	
Pseudo R^2	0.0335		0.837	

Table 2: Logit coefficients for demographics on worker types

α-workers, 7.4 thousand US$ for β-workers, and only 5.3 thousand US$ for γ-workers.

It turns out that a worker's type—a summary of the worker's pattern of employment transitions—is essentially orthogonal to the worker's demographic characteristics or industry. For every individual in our sample, we collect birth year, gender, race (white or non-white), education (some high school, high school, some college or college), State, and the 2-digit NAICS code for the industry where the individual is employed most frequently. We then run a multinomial logit regression on the probability of being a particular type of worker using demographics and industry as explanatory variables.

Table 2 reports the estimated regression coefficients—where the baseline outcome is white and the baseline demographics are male, white and some high school. The regression coefficients have the expected signs. For instance, an individual with more education is less likely to be an α-worker, a non-white individual is less likely to be an α-worker, etc.... These coefficients, though, are of little interest because the fit of the multinomial logit model is so poor. The pseudo R-squared of the regression is 3.35%, compared to a pseudo R-squared of 83.7% for a multinomial logit that also includes the individual's statistics on employment transitions. Similarly, the Tjur's R-squared—which measures the average difference between the predicted probability of type j if the worker's true type is j and if the worker's true type is different from j—is 4% for α-workers, 0.6% for β-workers, and 5% for γ-workers.

The finding that demographic characteristics and industry cannot predict a worker's type may come as a surprise. It is a well-established fact that the unemployment rate is very different for high school graduates and college graduates, for men and women, for young and old workers. And yet, education, gender and age cannot predict whether a worker is an α, β or γ. One way

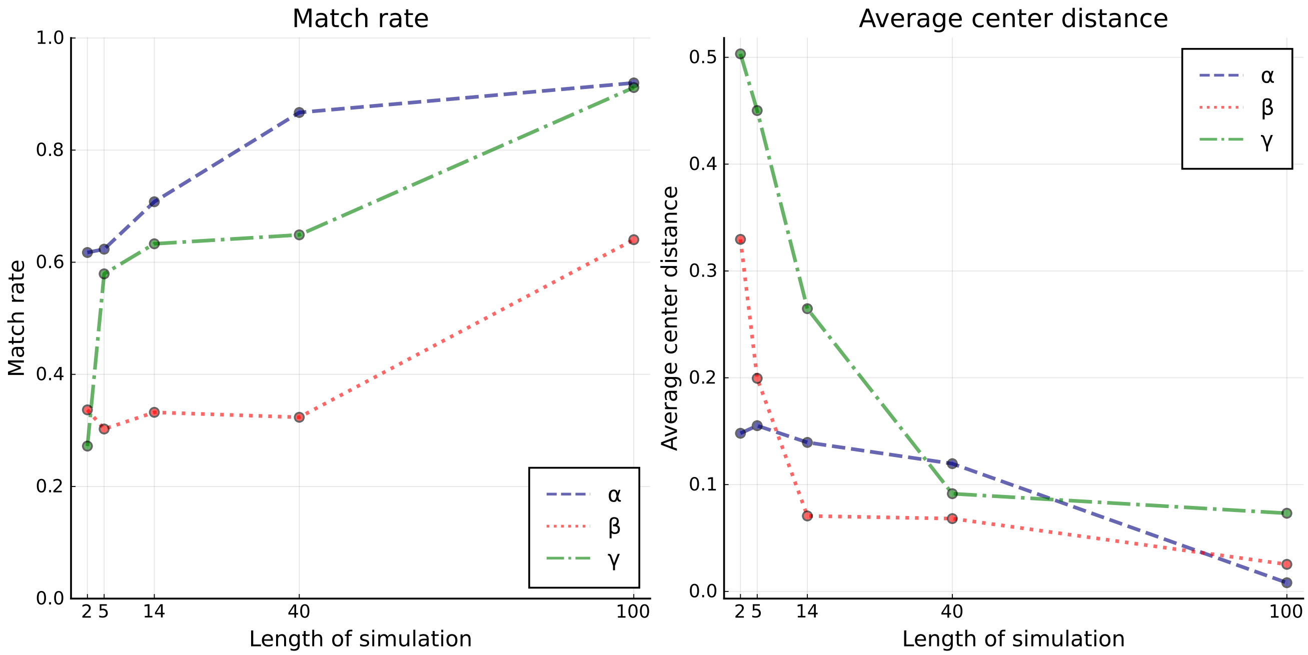

Figure 1: Left panel: the proportion of workers who are classified as type j that are truly of type j. Right panel: the difference between the average statistics of workers classified as type j and those of workers who are actually type j, measured in standard deviations.

to understand this finding is to relate it to Mincerian wage regressions. It is true that wages are very different for high school graduates and college graduates, for men and women, for young and old workers. And, yet, Mincer regressions show that these observable characteristics account for a very small fraction of the dispersion of wages in the cross-section. In other words, both for wages and for employment transitions, the differences in average outcomes across workers with different observables are small relative to the difference in outcomes within groups of workers who share the same observables.

2.4 Montecarlo analysis

While the asymptotic properties of the k-means algorithm are well understood (Bonhomme, Lamadon and Manresa 2019, 2021), it is useful to assess the performance of the algorithm in a finite sample. To this aim, we run several Montecarlo simulations. Using the theoretical model that will be presented in Section 3 and that is calibrated to reproduce the pattern of employment transitions of α, β and γ workers, we simulate individual histories. We then apply the k-means algorithm to the simulated individual histories and compare the resulting assignment of individuals to type with the actual individual types. Every simulation comprises half a million individual histories (the size of our sample from the LEHD) of different length: 2 years (the length of histories in the CPS), 5 years (the length of histories in the SIPP), 14 years (the length of histories from our sample of the LEHD), 40 years (about the longest histories one might observe), and 100 years.

In the left panel of Figure 1, we plot the fraction of workers classified as α, β or γ that are actually α, β or γ. As the length of the individual histories increases, the fraction of workers that are correctly classified by the algorithm increases. For 2-year long histories, 60% of workers classified as α's are actually α's, and 30% of the workers classified as β's and γ's are actually β's and γ's. For 100-year long histories, 90% of the workers classified as α's are actually α's, 90% of the workers classified as γ's are actually γ's, and 60% of the workers classified as β's are actually β's. For 14-year long histories, the fraction of workers who are correctly classified is 70% for αs, 60% for γs and 30% for β's. At the relevant length, the k-means algorithm does a good job at recognizing α and γ-workers. However, only 30% of workers classified as β's are actually β's, while the remaining 70% are either α's or γ's. Since the β-workers have a stochastic process for transitions that is intermediate between the process for α's and γ's, the fact that some lucky β's are classified as α's and some unlucky ones are classified as γ's is not surprising. It is also not too worrisome, as it involves misclassifying workers across types that are relatively similar.

In the right panel of Figure 1, we plot the distance, measured in standard deviations, between the average statistics of workers that are classified as α, β or γ and the average statistics of workers that are actually α, β or γ. As the length of individual histories increases, the distance quickly falls to zero for all types of workers. For 14-year histories, the distance between the average stats of actual and classified workers of type α is 5% of a standard deviation; it is essentially zero for β's; and it is about 30% of a standard deviation for γ's. We think that these distances are reasonably small.

3 Modeling Heterogeneity

In this section, we develop an equilibrium model of workers' transitions across employment states based on Menzio and Shi (2011) and Menzio, Telyukova and Visschers (2016). Firms spend resources to open vacancies and they advertise the terms of trade offered to workers hired to fill them. Workers spend time seeking vacancies and direct their search towards jobs offering particular terms of trade. In deciding where to search, workers trade off the probability of finding a job with the generosity of the terms of trade offered by the job, and the resolution of the trade-off depends on the worker's employment position. The quality of a particular firm-worker match is random and it is observed only after the match is consummated. If the quality is low enough, the worker moves into unemployment to find a new job as quickly as possible. If the quality of the match is in an intermediate range, the worker keeps searching while staying on the job. If the quality of the match is high enough, the worker stops searching altogether. The model generates endogenous workers' transitions between unemployment, employment and across employers. The transition probabilities differ across types of workers, as a worker's type affects his ability to search, the distribution of the quality of his match with different firms, and the speed at which he discovers the quality of a match.

3.1 Environment

The labor market is populated by a positive measure of workers and firms. Workers are ex-ante heterogeneous with respect to their type $i = 1, 2, \ldots I$. A worker of type i maximizes the present value of labor income discounted at the factor $\rho \in (0, 1)$. A worker of type i earns some income b_i when he is unemployed, and some income w_i when he is employed. The unemployment income b_i is a combination of unemployment benefits and value of leisure. The employment income w_i is determined by the worker's employment contract. The measure of workers of type i is $\mu_i \in [0, 1]$ and the total measure of workers is 1.

Firms are ex-ante homogeneous. Each firm maximizes the present value of profits, discounted by the factor ρ. Each firm operates a constant return to scale technology which turns the labor supply of a worker of type i into xy_iz units of output, where $x \in X \subset \mathbb{R}_+$ is a component of productivity that is common to all firm-worker pairs, $y_i \in Y \subset \mathbb{R}_+$ is a component that is common to all pairs of firms and workers of type i, and $z \in Z \subset R_+$ is a component that is specific to a particular firm-worker pair. The aggregate component of productivity x is time-varying, and it is the cause of aggregate labor market fluctuations. The type-specific component of productivity y_i is permanent, and it is the cause of differences in the average earnings of different types of workers. The match-specific component of productivity z is also permanent, and it is the cause of workers' job-to-job mobility. We refer to z as the *quality* of a firm-worker match. We assume that the quality of a firm-worker match is observed only after the match is consummated (i.e., matches are experience goods).

The labor market is organized in a continuum of submarkets indexed by $m = \{v, i\}$, where $v \in R$ denotes the lifetime income promised by firms to workers hired in m, and $i \in \{1, 2, ...I\}$ denotes the type of workers hired by firms in m. Associated with each submarket m, there is an endogenous vacancy-to-applicant ratio $\theta_i(v) \in \mathbb{R}_+$. We refer to $\theta_i(v)$ as the *tightness* of submarket m. If a worker applies for a job in m, he finds a vacancy with probability $p(\theta_i(v))$, where p is a strictly increasing, strictly concave function with $p(0) = 0$ and $p(\infty) = 1$. Similarly, if a firm opens a vacancy in m, it finds an applicant with probability $q(\theta_i(v))$, where q is a strictly decreasing function with $q(\theta) = p(\theta)/\theta$, $q(0) = 1$ and $q(\infty) = 0$.

At the beginning of each period, the state of the economy can be summarized by the aggregate component of productivity and by the distribution of workers across types and employment states. Formally, the state of the economy is given by $\psi \equiv \{x, u_i, n_i, g_i\}$, where $x \in X$ is the aggregate component of productivity, $u_i \in [0, 1]$ is the measure of workers of type i who are unemployed, $n_i \in [0, 1]$ is the measure of workers of type i who are employed in a match of unknown quality, $g_i : Z \to [0, 1]$ is a function such that $g_i(z)$ denotes the measure of workers of type i who are employed in a match of known quality z.

Each period consists of five stages: *entry-and-exit*, *learning*, *separation*, *search*, and *production*. At the *entry-and-exit* stage, a worker of type i exits the labor market with probability $1 - \chi$, with $\chi \in [0, 1]$. At the same time, a measure $(1 - \chi)\mu_i$ of workers of type i enters the labor market in the state of unemployment. Since the measure of workers of type i who exits the labor market is equal to the measure of workers entering the labor market, the measure of workers of type i in the economy remains constant over time.

At the *learning* stage, a worker of type i and a firm discover the quality z of their match with probability $\phi_i \in [0,1]$. The quality of the match is a random draw from a probability distribution function $f_i : Z \to [0,1]$ with a mean normalized to 1. At the *separation* stage, a match between a worker of type i and a firm breaks up with probability $d \in [\delta_i, 1]$ The probability d is specified by the employment contract regulating the relationship between the worker and the firm. The lower bound δ_i denotes the probability that the worker has to leave the match for exogenous reasons (e.g., firm closure or worker relocation).

At the *search* stage, a worker of type i gets the opportunity to search the labor market with a probability that depends on his employment status. If a worker is unemployed, he gets to search with probability $\lambda_u^i = 1$. If the worker is employed, he gets to search the market with probability $\lambda_e^i \in [0,1]$. If the worker became unemployed during the previous separation stage, he cannot search. Whenever the worker gets to search, he chooses in which submarket m to apply for a job. Simultaneously, firms choose how many vacancies to open in each submarket m at the unit cost $k_i > 0$.

Applicants and vacancies in submarket $m = \{v, i\}$ meet bilaterally according to the meeting probabilities $p(\theta_i(v))$ and $q(\theta_i(v))$. When a vacancy and an applicant of type i meet in m, the firm that owns the vacancy offers to the applicant a contract that is worth v in lifetime income. If the applicant accepts the offer, he becomes employed by the firm under the rules of the contract. If the applicant rejects the offer, which is an off-equilibrium event, he returns to his previous employment status. When a vacancy and an applicant of a type different from i meet in submarket m, the firm refuses to hire the applicant.[6]

At the *production* stage, a worker of type i who is unemployed receives an income equal to b_i units of output, a combination of unemployment benefits and value of leisure. A worker of type i who is employed in a match of unknown quality produces xy_i units of output and is paid some wage w_i, where w_i is specified by his employment contract. Similarly, a worker of type i who is employed in a match of known quality z produces xy_iz units of output and is paid some wage w_i. After production takes place, next period's aggregate component of productivity, $\hat{x}$, is drawn from the probability density function $h : X \times X \to \mathbb{R}_+$ with $h(\hat{x}, x)$ denoting the probability density of $\hat{x}$ conditional on x.

We assume that the contracts offered by firms to workers are *bilaterally efficient*, in the sense that they maximize the joint present value of income of the firm and the worker. As discussed in Menzio and Shi (2011), this assumption is consistent with several contractual environments. Consider two cases. In the first case, a contract can specify the worker's wage, the worker's search strategy on the job (i.e. in which submarket to search) and the worker's quitting strategy (i.e. when to move into unemployment) contingent on the history of the match and the economy. In this case, the contract space is rich enough to independently control the

[6]We assume that a worker knows his own type and so does the market. The second part of the assumption may appear unrealistic, but it does greatly simplify the analysis. Specifically, the assumption allows us to abstract from issues of signaling—the worker distorting his behavior so as to convince the market that his type is better than what it actually is—as well as from issues of inference–the firms having to assess the probability distribution of a worker's type by examining his employment history and performance on the job.

allocative decisions of the match and the distribution of the value of the match between the firm and the worker. Given this contractual environment, the firm finds it optimal to offer a contract such that the allocative decisions maximize the joint income of the match, and such that the wages provide the worker with the lifetime income v. In the second case, a contract can specify a sign-on transfer and then a wage contingent on the history of the match and the economy. The worker is then free to follow his preferred search and quitting strategy. In this case, the firm finds it optimal to offer a contract such that the worker is the residual claimant of output (and, hence, makes allocative decisions to maximize the joint income of the match) and a (possibly negative) transfer such that the worker's lifetime income is v.

Before turning to the definition of equilibrium, it is useful to briefly motivate our approach to modelling heterogeneity in the workers' employment transitions. We assume that a worker's type i affects k_i and λ_e^i in order to capture the fact that types are heterogeneous with respect to the speed at which they move from unemployment to employment and across different employers.[7] We assume that a worker's type affects f_i, ϕ_i and δ_i in order to capture the fact that types are heterogeneous with respect to the distribution of job durations. Lastly, we assume that a worker's type affects y_i in order to capture type heterogeneity with respect to average earnings. Since unemployment benefits are related to earnings, we also let b_i be type-specific.

3.2 Equilibrium

To formally define an equilibrium, we need to introduce a few additional pieces of notation. Let $U_i(\psi)$ denote the lifetime income for a worker of type i who is unemployed at the beginning of the production stage. Let $\tilde{V}_i(\psi)$ denote the sum of the lifetime income for a firm and a worker of type i who, at the beginning of the production stage, are in a match of unknown quality. Let $V_i(z,\psi)$ denote the sum of the lifetime income for a firm and a worker of type i who, at the beginning of the production stage, are in a match of known quality z. Lastly, let $\theta_i(v,\psi)$ denote the equilibrium tightness of submarket $m = \{v, i\}$.

The value $U_i(\psi)$ of unemployment for a worker of type i is given by

$$U_i(\psi) = b_i + \rho\chi\mathbb{E}_{\hat{\psi}}\left[U_i(\hat{\psi}) + \lambda_u^i \max_v \left\{p(\theta_i(v,\hat{\psi}))\left(v - U_i(\hat{\psi})\right)\right\}\right]. \tag{3.1}$$

In the current period, the worker's income is b_i. In the next period, the worker finds a job with probability $\lambda_u^i p(\theta_i(v,\hat{\psi}))$. In this case, the worker's continuation value is v. The worker does not find a job with probability $1 - \lambda_u^i p(\theta_i(v,\hat{\psi}))$. In this case, the worker's continuation value is $U_i(\hat{\psi})$.

The joint value $V_i(z,\psi)$ of a match of quality z between a firm and a worker of type i is

[7] We generate differences in the unemployment duration of different types of workers by letting the vacancy cost k_i depend on the type of worker that the firm is seeking, while keeping λ_u^i equal to 1 for all types of workers. Alternatively, we could let the probability λ_u^i that a worker can search the labor market when unemployed depend on the worker's type, and kept $k_i = k$ for all types of workers. It is easy to verify that the two approaches lead to the same equilibrium outcomes.

given by

$$
\begin{aligned}
&V_i(z,\psi) = xy_i z \\
&+\rho\chi\mathbb{E}_{\hat{\psi}}\left[\max_{d\in[\delta,1]} dU_i(\hat{\psi}) + (1-d)\left[V_i(z,\hat{\psi}) + \lambda_e^i \max_v \left\{p(\theta_i(v,\hat{\psi}))\left(v - V_i(z,\hat{\psi})\right)\right\}\right]\right]
\end{aligned}
\tag{3.2}
$$

In the current period, the sum of the worker's income and firm's profit is xy_iz, the output of the match. In the next period, the worker moves into unemployment with probability d. In this case, the worker's continuation value is $U_i(\hat{\psi})$ and the firm's continuation value is 0. The worker moves from the current job to a new job with probability $(1-d)\lambda_e^i p(\theta_i(v,\hat{\psi}))$. In this case, the worker's continuation value is v and the firm's continuation value is 0. The worker and the firm remain together with probability $(1-d)(1-\lambda_e^i p(\theta_i(v,\hat{\psi})))$. In this case, the firm's and worker's joint continuation value is $V_i(z,\hat{\psi})$. Note that, since employment contracts are bilaterally efficient, d and v are chosen so as to maximize the joint value of the match.

The joint value $\tilde{V}_i(\psi)$ of a match of unknown quality between a firm and a worker of type i is given by

$$
\begin{aligned}
&\tilde{V}_i(\psi) = xy_i \\
&+\rho\chi\phi_i\mathbb{E}_{z,\hat{\psi}}\left[\max_{d\in[\delta,1]} dU_i(\hat{\psi}) + (1-d)\left[V_i(z,\hat{\psi}) + \lambda_e^i \max_v \left\{p(\theta_i(v,\hat{\psi}))\left(v - V_i(z,\hat{\psi})\right)\right\}\right]\right] \\
&+\rho\chi(1-\phi_i)\mathbb{E}_{\hat{\psi}}\left[\max_{d\in[\delta,1]} dU_i(\hat{\psi}) + (1-d)\left[\tilde{V}_i(\hat{\psi}) + \lambda_e^i \max_v \left\{p(\theta_i(v,\hat{\psi}))\left(v - \tilde{V}_i(\hat{\psi})\right)\right\}\right]\right]
\end{aligned}
\tag{3.3}
$$

In the current period, the expected output of the match is xy_i. In the next period, the firm and the worker learn the quality z of their match with probability ϕ_i. The worker leaves the match for unemployment with probability d. In this case, the joint continuation value is $U_i(\hat{\psi})$. The worker searches on-the-job and finds a new job with probability $(1-d)\lambda_e^i p(\theta_i(v,\hat{\psi}))$. In this case, the joint continuation value is v. The worker and the firm remain together with probability $(1-d)(1-\lambda_e^i p(\theta_i(v,\hat{\psi})))$. In this case, the joint continuation value is $V_i(z,\hat{\psi})$. The firm and the worker do not learn the quality of their match with probability $1-\phi_i$. Since employment contracts are bilaterally efficient, the choice of d and v is contingent on whether the quality of the match is observed or not and, if it is, on the realization of z.

The tightness $\theta_i(v,\psi)$ of submarket $m = \{v, i\}$ is such that

$$
k_i \geq q(\theta_i(v,\psi))(\tilde{V}_i(\psi) - v),
\tag{3.4}
$$

and $\theta_i(v) \geq 0$, with the two inequalities holding with complementary slackness. The left-hand side of (3.4) is the cost to a firm from opening a vacancy in submarket m. The right-hand side is the benefit to the firm from opening a vacancy in submarket m. The benefit is the probability that the firm fills its vacancy, $q(\theta_i(v))$, times the firm's value from filling a vacancy, $\tilde{V}_i(\psi) - v$, i.e. the joint value of a match between the firm and a worker of type i net of the lifetime utility promised by the firm to the worker. Condition (3.4) then states that the cost and benefit of a vacancy in submarket m must be equal if the vacancy-to-applicant ratio is strictly positive.

And the vacancy-to-applicant ratio must be equal to zero if the cost of a vacancy in submarket m is strictly greater than the benefit. In submarkets with some applicants, the condition guarantees that the tightness is consistent with firm's profit maximization. In submarkets without applicants, the condition pins down the agents' expectations about the tightness.

We now turn to characterizing the solution to the search and separation problems in (3.1), (3.2), (3.3). The search problem for a worker of type i who currently is an employment state with value v_0 is given by

$$D_i(v_0, \psi) = \max_{v} p(\theta_i(v, \psi))(v - v_0). \tag{3.5}$$

For any v such that $\theta_i(v, \psi) > 0$, (3.4) implies that v is equal to $\tilde{V}_i(\psi) - k_i/q(\theta_i(v, \psi))$ and, hence, the objective function in (3.5) is equal to $p(\theta_i(v, \psi))(\tilde{V}_i(\psi) - v_0) - k\theta_i(v, \psi)$. For any v such that $\theta_i(v, \psi) = 0$, $p(\theta_i(v, \psi)) = 0$ and, hence, the objective function in (3.5) is also equal to zero or, equivalently, to $p(\theta_i(v, \psi))(\tilde{V}_i(\psi) - v_0) - k\theta_i(v, \psi)$.

The above observations allow us to rewrite the search problem in (3.5) as

$$D_i(v_0, \psi) = \max_{v} -k_i\theta_i(v, \psi) + p(\theta_i(v, \psi))(\tilde{V}_i(\psi) - v_0). \tag{3.6}$$

Notice that, for all $\theta \geq 0$, there exists a v such that $\theta_i(v, \psi) = \theta$. Thus, by changing the choice variable from v to θ in (3.6), we do not enlarge the choice set. Conversely, for all v, there exists a $\theta \geq 0$ such that $\theta = \theta_i(v, \psi)$. Thus, by changing the choice variable from v to θ in (3.6), we do not shrink the choice set. Since the choice set is the same whether the worker chooses v or θ, we can rewrite (3.6) as

$$D_i(v_0, \psi) = \max_{\theta \geq 0} -k_i\theta + p(\theta)(\tilde{V}_i(\psi) - v_0). \tag{3.7}$$

In words, the worker chooses the tightness θ of the submarket in which to apply for a job so as to maximize the probability of meeting a firm, $p(\theta)$, times the difference between the joint value of the match with the firm and the value of his current employment state, $\tilde{V}_i(\psi) - v_0$, net of the firm's cost of opening θ vacancies, $k_i\theta$.

The solution to the worker's search problem in (3.7) satisfies the following necessary and sufficient condition for optimality

$$k_i \geq p'(\theta)(\tilde{V}_i(\psi) - v_0), \tag{3.8}$$

and $\theta \geq 0$, where the two inequalities hold with complementary slackness. In words, (3.8) states that, if the worker searches in a submarket with a strictly positive tightness, the cost, kk_i, of searching in a submarket with a marginally higher tightness must be equal to the benefit, $p'(\theta)(\tilde{V}_i(\psi) - v_0)$. If the worker searches in a submarket with zero tightness, the marginal cost must be greater of equal to the marginal benefit. We denote as $\theta^*_{i,u}(\psi)$ the optimal search strategy for a worker who is unemployed. That is, $\theta^*_{i,u}(\psi)$ is the solution to (3.8) for $v_0 = U_i(\psi)$. We denote as $\theta^*_{i,e}(z, \psi)$ the optimal search strategy for a worker who is employed in a match of known quality z. That is, $\theta^*_{i,e}(z, \psi)$ is the solution to (3.8) for $v_0 = V_i(z, \psi)$. Since $V_i(z, \psi)$ is strictly increasing in z, $\theta^*_{i,e}(z, \psi)$ is strictly decreasing for all $z < Q_i(\psi)$ and zero for all

$z \geq Q_i(\psi)$, where $Q_i(\psi)$ is defined as the quality of a match that has the same joint value as a match of unknown quality, i.e. $V_i(Q_i(\psi), \psi) = \tilde{V}_i(\psi)$. Obviously, a worker employed in a match of unknown quality finds it optimal to search in a submarket with zero tightness.

Next, we turn to the characterization of the separation problems in (3.2) and (3.3). The optimal separation probability for a firm and a worker of type i who are in a match with some joint value v_0 is determined by the sign of the following inequality

$$U_i(\psi) \lesseqgtr v_0 + \lambda_e^i D_i(v_0, \psi). \tag{3.9}$$

The left-hand side is the firm's and worker's joint value of breaking up at the separation stage. The right-hand side is the firm's and worker's joint value of remaining together at the separation stage. If the left-hand side is greater than the right-hand side, then the optimal separation probability is equal to 1. Otherwise, it is equal to δ_i. We denote as $d_i^*(z, \psi)$ the optimal separation probability for a firm and a worker in a match of known quality z. That is, $d_i^*(z, \psi)$ denotes the optimal separation probability for $v_0 = V_i(z, \psi)$. Since $V_i(z, \psi)$ is strictly increasing in z, there exists a reservation quality $R_i(\psi)$ such that $d_i^*(z, \psi) = 1$ for all $z < R_i(\psi)$ and $d_i^*(z, \psi) = \delta_i$ for all $z \geq R_i(\psi)$. Similarly, we denote as $\tilde{d}^*(\psi)$ the optimal separation probability for a firm and a worker in a match of unknown quality.

We now have a complete characterization of the workers' transitions across employment states. An unemployed worker finds a job with probability $\lambda_u^i p(\theta_i(v, \psi))$. As long as the worker does not observe the quality of his match, he moves into unemployment with probability $\tilde{d}^*(\psi)$ and he does not search for a better job—in the sense that he searches in a submarket with zero tightness. Once the worker observes z, he moves into unemployment with probability 1 if $z < R_i(\psi)$. If $z \in [R_i(\psi), Q_i(\psi))$, the worker moves into unemployment only for exogenous reasons, and actively searches for a better job—in the sense that he searches in a submarket with positive tightness. If $z \geq Q_i(\psi)$, the worker moves into unemployment only for exogenous reasons and does not search for a better job.

To conclude the definition of equilibrium, we need the laws of motion for the distribution of workers across employment states. The law of motion for the measure u_i of workers of type i who are unemployed is given by

$$\begin{aligned} \hat{u}_i &= (u_i \chi + (1-\chi)\mu_i)\left(1 - \lambda_u^i(\theta_{i,u}^*(\psi))\right) \\ &\quad + \textstyle\sum_z \left[(g_i(z)\chi + n_i \chi \phi_i f_i(z)) d_i^*(z, \psi)\right] + n_i \chi (1-\phi_i) \tilde{d}_i^*(\psi). \end{aligned} \tag{3.10}$$

The left-hand side of (3.10) is the measure of unemployed workers at the beginning of next period. The first term on the right-hand side is the measure of workers who are unemployed at the beginning of the current period and do not find a job. The second term sums the measure of workers who are employed in a match of known quality at the beginning of the current period and become unemployed at the separation stage with the measure of workers who are employed in a match of unknown quality at the beginning of the current period, discover the quality of their match at the learning stage, and become unemployed at the separation stage. The last term is the measure of workers who are employed in a match of unknown quality at the beginning of the current period, do not discover the quality of their match at the learning stage, and become unemployed at the separation stage.

The law of motion for the measure n_i of workers of type i who are employed in a match of unknown quality is given by

$$\begin{aligned} \hat{n}_i \quad &= n_i\chi(1-\phi_i)(1-\tilde{d}_i^*(\psi)) + (u_i\chi + (1-\chi)\mu_i)\, \lambda_u^i(\theta_{i,u}^*(\psi)) \\ &+ \textstyle\sum_z (g_i(z)\chi + n_i\chi\phi_i f_i(z))(1-d_i^*(z,\psi))\lambda_e^i(\theta_{i,e}^*(z,\psi)). \end{aligned} \tag{3.11}$$

The left-hand side of (3.11) is the measure of workers employed in a match of unknown quality at the beginning of next period. The first term on the right-hand side of (3.11) is the measure of workers who are employed in a match of unknown quality at the beginning of the current period, do not discover the quality of their match at the learning stage, and remain on their job. The second term is the measure of workers who are unemployed at the beginning of the current period and find a job at the search stage. The last term is the measure of workers who are employed at the beginning of the current period and move to a new job during the search stage.

The law of motion for the measure $g_i(z)$ of workers of type i who are employed in a match of known quality z is given by

$$\begin{aligned} \hat{g}_i(z) \quad &= g_i(z)\chi(1-d_i^*(z,\psi))(1-\lambda_e^i p(\theta_{i,e}^*(z,\psi))) \\ &+ n_i\chi\phi_i f_i(z)(1-d_i^*(z,\psi))(1-\lambda_e^i p(\theta_{i,e}^*(z,\psi))). \end{aligned} \tag{3.12}$$

The left-hand side is the measure of workers employed in a match of quality z at the beginning of next period. The first term on the right-hand side is the measure of workers who are employed in a match of quality z at the beginning of the period and remain on their job. The second term is the measure of workers who are employed in a match of unknown quality at the beginning of the period, discover that the quality of their match is z during the learning stage, and remain on their job.

We are now in the position to define a Recursive Equilibrium.

Definition 1. *A Recursive Equilibrium (RE) is given by value functions $\{U_i, \tilde{V}_i, V_i\}$, policy functions $\{d_i^*, \tilde{d}_i^*, \theta_{i,u}^*, \theta_{i,e}^*\}$, and a transition probability function $\Phi(\hat{\psi}|\psi)$ for the aggregate state of the economy such that: (i) $\{U_i, \tilde{V}_i, V_i\}$ satisfy (3.1), (3.2) and (3.3); (ii) $\{d_i^*, \tilde{d}_i^*, \theta_{i,u}^*, \theta_{i,e}^*\}$ satisfy the optimality conditions (3.8) and (3.9); Φ is consistent with the laws of motion (3.10), (3.11) and (3.12) for $\{\hat{u}_i, \hat{n}_i, \hat{g}_i\}$ and with the probability distribution for $\hat{x}$.*

We can also define a Block Recursive Equilibrium.

Definition 2. *A Block Recursive Equilibrium (BRE) is a RE in which the value and policy functions depend on the aggregate state of the economy $\psi = \{x, u_i, n_i, g_i\}$ only through the state of productivity x and not through the distribution of workers across employment states $\{u_i, n_i, g_i\}$.*

As established in Menzio and Shi (2011), there exists a unique BRE and no other RE. Moreover, the BRE is efficient in the sense that it decentralizes the solution of the problem of a utilitarian social planner. Let us provide some intuition for the existence and uniqueness of a BRE. A BRE exists because search is directed, and the search and production processes feature constant returns to scale. Consider the equilibrium condition (3.4) for the tightness of submarket $m = \{v, i\}$. Since the production process features constant returns to scale, the value to the firm from filling a vacancy in submarket m depends on the aggregate productivity x and on the promised value v and not on the distribution of workers $\{u_i, n_i, g_i\}$. Since the search process is directed, the probability that a worker contacted in submarket m is willing to accept the value offered by the firm is always equal to one. Since the search process features constant returns to scale, the probability of contacting a worker in submarket m only depends on the tightness of the submarket and not on the measure of workers searching there. The above observations imply that the tightness of submarket m depends on x but not on the distribution of workers $\{u_i, n_i, g_i\}$. In turn, this implies that the value functions—as well as the associated policy functions—are all independent of the endogenous distribution of workers across employment states. The uniqueness of the BRE follows from the fact that the value functions can be combined in a single operator and this operator is a contraction.

4 Calibration

In this section, we carry out the second step of the two-stage Group Fixed Effect (GFE) method. Specifically, we are going to calibrate the type-specific parameters of our model of employment transitions by matching type-specific empirical moments that we computed using the k-means discretization algorithm. In Section 4.1, we motivate our choice of empirical moments used to calibrate the parameters. In Section 4.2, we present and discuss the calibration outcomes and the key differences between types with respect to the calibrated parameter values.

4.1 Calibration strategy

We calibrate the parameters of the model by matching the moments generated by the model at its non-stochastic steady-state with the analogous moments observed in the data over the pre-Great Recession period. The non-stochastic steady state is defined as the steady state associated with a version of the model in which the aggregate component of productivity x is kept constant and equal to the unconditional mean of the stochastic process $h(\hat{x}|x)$. The period before the Great Recession is defined as the period between 1997 and 2008.

Let us begin by reviewing the parameters that need to be calibrated. The parameters describing the production process are: (i) the unconditional mean x^* of the aggregate component of productivity, which we normalize to 1; (ii) the component of productivity y_i that is specific to a worker of type i; (iii) the distribution f_i of the component of productivity z that is specific to the match between a particular worker of type i and a particular firm; (iv) the probability ϕ_i with which a worker of type i and a firm discover the quality of their match. We specialize f_i to be a Weibull distribution with shape parameter ω_i and scale parameter σ_i that is appropriately relocated so as to have a mean of one. The Weibull distribution is flexible and, depending on the parameter ω_i, its shape can resemble an exponential, a log-normal, a normal, or a left-skewed distribution.[8]

The parameters describing the search process are: (i) the probability λ_u^i that a worker of type i can search the labor market when unemployed; (ii) the probability λ_e^i that a worker of type i can search the labor market when employed; (iii) the probability $p(\theta)$ that an applicant meets a vacancy as a function of the tightness θ; (iv) the probability δ_i that the match between a worker of type i and a firm breaks up for exogenous reasons. Following much of the labor-search literature, we specialize $p(\theta)$ to have the form $p(\theta) = \min\{\theta^\gamma, 1\}$, where γ denotes the elasticity of the job-finding probability with respect to tightness and is set to 0.5.[9]

The parameters describing the population and the preferences of workers are: (i) the measure μ_i of workers of type i; (ii) the probability χ that a worker does not exit the labor market; (iii) the workers' discount factor ρ; (iv) the sum b_i of the unemployment income and the value of leisure for workers of type i. We specialize b_i to be of the form $\zeta + r\mathbb{E}[xy_iz]$, where ζ denotes the value of leisure and r denotes the fraction of the average productivity $\mathbb{E}[xy_iz]$ that is replaced by unemployment benefits.

Let us now turn to the calibration strategy.[10] Based on the clustering analysis of Section 2, we calibrate the model to have three types of workers: α, β and γ. We calibrate the measure μ_i of workers of type i so as to match the empirical distribution of workers across types, i.e. we set $\mu_\alpha = 0.57$, $\mu_\beta = 0.26$ and $\mu_\gamma = 0.17$. We set the length of a period to be equal to one month. We set the discount factor ρ to be 0.996, which implies an annual interest rate of 5%. We set the probability χ that a worker remains in the labor market to 0.996, which implies an average work-life of about 20 years.

We calibrate the cost k_i of opening a vacancy to hire workers of type i so as to match the average UE rate of workers of type i, i.e. a UE rate of about 30% for α-workers, 15% for β-workers, and 10% for γ-workers.[11] Similarly, we calibrate the probability δ_i that a worker of type i moves into unemployment for exogenous reasons so as to match the average unemployment

[8]Menzio, Telyukova and Visschers (2016) and Martellini, Menzio and Visschers (2021) use models that are very similar to ours and, at the calibration stage, they also specialize the match-quality distribution to be Weibull.

[9]Assuming that an applicant meets a vacancy with probability $p(\theta) = \min\{\theta^{0.5}, 1\}$ is equivalent to assuming that applicants and vacancies come together via a Cobb-Douglas matching function with elasticity of 0.5 with respect to both applicants and vacancies. Such a specification of the matching function is ubiquitous in quantitative applications of labor-search models.

[10]In order to clearly explain our calibration strategy, we "pretend" that each parameter of the model is chosen to reproduce one particular moment or group of moments in the data. In reality, the calibration algorithm simultaneously chooses the parameter values to minimize the distance with respect to all of the targeted moments.

[11]We measure the average UE rate for workers of type i as the UE rate that best fits the distribution of their unemployment spells. Note that, in the data, the type-specific UE rate is slightly negatively correlated with the duration of unemployment. In contrast, in the model, the type-specific UE rate is constant. In order to capture the decline in the type-specific UE rate, we could have generalized the model to allow for lambdai(u) to fall with the duration of an unemployment spell. We decided against this generalization in order to keep the model simpler.

rate of workers of type i in the period preceding the Great Recession, i.e. an unemployment rate of 4.2% for α-workers, 12.5% for β-workers, and 28.8% for γ-workers. The choice of calibration targets is natural, since k_i affects the tightness function $\theta_i(v)$ and, in turn, the UE rate for workers of type i. Having matched the UE rate, δ_i affects the EU rate for workers of type i and, in turn, their unemployment rate.

The scale σ_i of the match quality distribution f_i is chosen so as to reproduce the fraction of matches between a firm and a worker of type i that terminate before reaching 2 years of tenure, i.e. about 50% for α-workers, 60% for β-workers, and 85% for γ-workers. The choice of the calibration target is easy to understand, since σ_i affects the probability that the quality z of a match between a firm and a worker is smaller than R_i—which induces the worker to move into unemployment to search for a better match—and the probability that the z is between between R_i and Q_i—which induces the worker to search for a better match on the job. In turn, these probabilities affect the probability that a match between a firm and worker terminates before reaching 2 years of tenure.

The probability λ_u^i that an unemployed worker of type i gets to search the labor market is normalized to 1 for all worker types. The probability λ_e^i that an employed worker of type i gets to search the labor market is chosen so as to reproduce the fraction of matches between a firm and a worker of type i that last less than 2 years and terminate with the worker moving directly to another employer. For α-workers, this fraction is 21% (about one-half of matches that last no more than 24 months). For β-workers, it is 18% (about one-third of matches that last no more than 24 months). For γ-workers, it is 22% (or about one-fourth of matches that last no more than 24 months).

The shape ω_i of the match-quality distribution f_i and the probability ϕ_i with which a firm-worker pair discovers the quality of their match are chosen to reproduce the whole shape of the tenure distribution. Specifically, the parameters are chosen so as to minimize the distance between the model and the data with respect to: (i) the fraction of firm-worker matches that terminate before exceeding 3 months of tenure, 12 months of tenure, and 24 months of tenure; (ii) the fraction of firm-worker matches that terminate with the worker moving to another employer before exceeding 3, 12 and 24 months of tenure; (iii) the fraction of firm-worker matches that terminate with the worker moving into unemployment before exceeding 3, 12 and 24 months of tenure. That is, ω_i and ϕ_i are chosen to fit the shape of the tenure distribution (unconditional, and conditional on the type of termination). The shape of the tenure distribution is quite different for different types. For instance, for α-workers, the fraction of matches ending within the first 3 months is lower than the fraction of matches ending between 13 and 24 months. For γ-workers, the fraction of matches ending within the first 3 months is much higher than the fraction of matches ending between 13 and 24 months. Our choice of these calibration targets for ω_i and ϕ_i is natural, as ϕ_i determines how quickly low-quality matches are identified, and ω_i determines the shape of the match-quality distribution (and, hence, the incentives to searching for a better match).

We normalize the component of productivity y_α that is specific to α-workers to 1. We choose the component of productivity y_i for $i = \{\beta, \gamma\}$ so that the model-generated ratio between the

Parameter	Value	Description
β	0.996	discount factor
b_i	(0.676, 0.533, 0.434)	flow unemployment income
y_i	(1, 0.623, 0.459)	type-specific productivity
α_i	(4.515, 3.941, 0.640)	shape of f_i
σ_i	(0.058, 0.143, 0.082)	standard deviation of f_i
ϕ_i	(0.307, 0.233, 0.229)	probability match quality is discovered
λ_e^i	(0.151, 0.493, 0.641)	probability an employed worker searches
λ_u^i	1	probability an unemployed worker searches
δ_i	(0.006, 0.009, 0.005)	exogenous separation probabilities
k_i	(2.808, 4.437, 2.605)	vacancy posting cost
γ	0.5	elasticity of job-finding rate wrt tightness
$1-\chi$	0.004	exogenous labor market exit probability

Table 3: Model parameters

average productivity among employed workers of type i and the average productivity among employed workers of type α is equal to the empirical ratio between the average earnings of employed workers of type i and the average earnings of employed workers of type α. The attentive reader may have noticed that in the calibration of y_i we compare productivity in the model with earnings in the data. We do so because computing productivity is easier than computing wages and, for the most common specification of the wage process (e.g., the wage is set equal to some fraction of the worker's productivity as in Bagger et al. 2014 or Menzio, Telyukova and Visschers 2016), the difference between productivity and wage turns out to be negligible.

Lastly, we need to calibrate the parameters associated with the unemployment income. Shimer (2005) argues that unemployment income should be set to 40% of average productivity, as this is the typical replacement rate in the US. Hagedorn and Manovskii (2008) point out that unemployment income should also include the value of leisure. Hall and Milgrom (2008) argue that, on average, the ratio between unemployment income (unemployment benefits plus value of leisure) is about 65% of employment income. Based on these observations, we choose the replacement ratio r of unemployment benefits for workers of type i to be equal to 40% of the average productivity among employed workers of type i. We then choose the value of leisure ζ so that the ratio between unemployment income and labor productivity is, on average, equal to 65%.

4.2 Calibration outcomes

Table 3 reports the calibrated value of the parameters of the model. It is useful to highlight the major differences between types with respect to the calibrated parameter values. Figure 2 plots the calibrated distribution of the match-specific quality for workers of type i, together with R_i—the cutoff below which workers find it optimal to move into unemployment—and Q_i—the cutoff above which workers find it optimal to stop searching for a better match. For α-workers, the calibrated distribution is a Weibull with shape 4.5 and scale 0.25. Such distribution

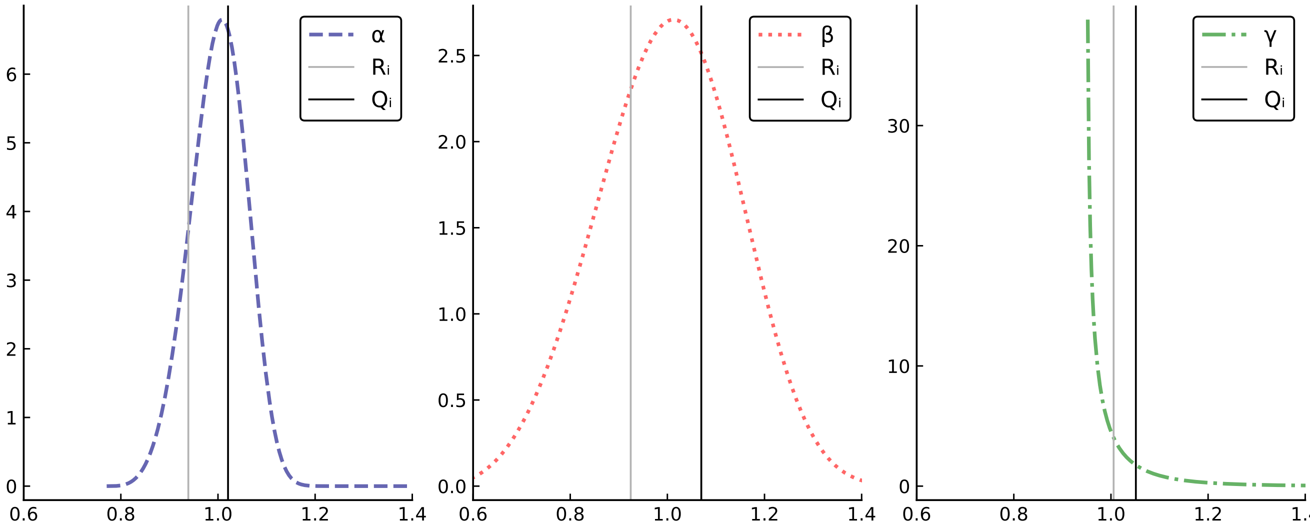

Figure 2: Match quality distributions $f_i(z)$

is approximately normal, with a mean of 1, a standard deviation of 0.06, a skewness of -0.17, and a 90-50 percentile ratio equal to 90% of the 50-10 percentile ratio. For β-workers, the calibrated distribution is a Weibull with shape 3.9 and scale 0.55. Such distribution is approximately normal, with a mean of 1, a standard deviation of 0.14, a skewness of -0.07, and a 90-50 percentile ratio equal to 93% of the 50-10 percentile ratio. For γ-workers, the calibrated distribution is a Weibull with shape 0.64 and scale 0.04. Such distribution is approximately exponential, with a mean of 1, a standard derivation of 0.08, a skewness of 4.12, and a 90-50 percentile ratio that is 6 times larger than the 50-10 percentile ratio.

The calibrated distribution of match qualities is not the same for all types because different types feature a different distribution of job durations, as shown in Table 4. Workers of type α have a 50% probability of remaining on a job for more than 2 years, and a 28% probability of leaving a job within the first year. In order to reproduce these facts, the calibrated match-quality distribution has relatively little variance and relatively small tails. The small left tail implies that the fraction of matches below R_i is small. The small right tail implies that the return to searching for better matches is low and, hence, Q_i is close to R_i. Workers of type β have a 40% probability of remaining on a job for more than 2 years, and a 38% probability of leaving a job within the first year. In order to reproduce these facts, the calibrated match-quality distribution has higher variance than for α-workers. Workers of type γ have only a 15% probability of remaining on a job for more than 2 years, and a striking 65% probability of leaving a job within 1 year. In order to reproduce these facts, the calibrated match-quality distribution is right-skewed and has a long right tail. The long right tail of the distribution gives workers the incentive to continue searching for a better match—even when they are employed in a match with a quality at the top 20% of the distribution. The left-skewness of the distribution implies that a large fraction of matches is below R_i and, hence, are terminated as soon as their quality is observed.

Another difference between types are the parameters describing the search process. For α-workers, the cost of maintaining a vacancy is equal to 2.8 units of output, the probability of

	α-workers					
	Data			**Model**		
	Total	Ends in E	Ends in U	Total	Ends in E	Ends in U
<1Q	0.122	0.077	0.045	0.099	0.088	0.010
1Q-4Q	0.162	0.087	0.075	0.193	0.111	0.082
5Q-8Q	0.207	0.101	0.105	0.138	0.050	0.087
>8Q	0.510	0.258	0.252	0.571	0.394	0.177
	β-workers					
	Data			**Model**		
	Total	Ends in E	Ends in U	Total	Ends in E	Ends in U
<1Q	0.180	0.141	0.040	0.149	0.138	0.011
1Q-4Q	0.209	0.143	0.066	0.302	0.206	0.096
5Q-8Q	0.219	0.147	0.072	0.154	0.065	0.089
>8Q	0.392	0.289	0.103	0.395	0.298	0.098
	γ-workers					
	Data			**Model**		
	Total	Ends in E	Ends in U	Total	Ends in E	Ends in U
<1Q	0.354	0.279	0.075	0.308	0.305	0.003
1Q-4Q	0.309	0.219	0.090	0.435	0.404	0.031
5Q-8Q	0.192	0.132	0.060	0.081	0.051	0.031
>8Q	0.144	0.103	0.041	0.176	0.137	0.039

Table 4: Employment duration moments

searching on the job is 15%, and the probability of losing a job for exogenous reasons is 0.6%. These parameters—together with the others—imply an unemployment rate of 4.2%, a UE rate of 30%, a EU rate of 0.9% and an EE rate of 0.6% per month. For β-workers, the cost of maintaining a vacancy is equal to 4.4 units of output, the probability of searching on the job is 49%, and the probability of losing a job for exogenous reasons is 0.9%. These parameters imply an unemployment rate of 12.4%, a UE rate of 15%, an EU rate of 1.7% and an EE rate of 0.8% per month. For γ-workers, the cost of maintaining a vacancy is equal to 2.6 units of output, the probability of searching on the job is 64%, and the probability of losing a job for exogenous reasons is 0.5%. These parameters imply an unemployment rate of 29.7%, a UE rate of 10%, an EU rate of 3.8%, and an EE rate of 0.4% per month.

Lastly, we want to point out the differences across types in productivity and unemployment income. For α-workers, the baseline productivity is 1 unit of output and the unemployment income is equal to 0.67 units of output—which is approximately equal to 64% of their average labor productivity. For β-workers, the baseline productivity is 0.62 units of output and the unemployment income is 0.53 units of output—which is approximately equal to 75% of their average labor productivity. For γ-workers, the baseline productivity is 0.46 units of output and the unemployment income is 0.43—which is approximately equal to 90% of their average labor productivity. These calibration outcomes reproduce the difference between types with respect to their labor earnings, a replacement ratio of unemployment benefits that is equal to 40% of average productivity, and a value of leisure that is common for all types and equal, on average, to 25% of average productivity.

The calibration outcomes provide a structural interpretation to the empirical differences in the pattern of employment transitions of different types of workers. Workers of type α have high baseline productivity and, hence, face large gains from trading in the labor market—which leads them to have a high UE rate. Workers of type α have a similar productivity when matched with different firms—which results in having a high probability of remaining on a job for a long period of time. In contrast, workers of type γ have low productivity and, hence, small gains from trading in the labor market—which leads them to have a low UE rate. Workers of type γ have very different productivity when matched with different firms and, in particular, they are much more productive when matched with a small subset of firms. This results in γ-workers having a low probability of remaining on a job for a long period of time. Overall, the search process of α-workers—which entails finding any job—is fast. The search process of γ-workers—which entails finding one of the rare good jobs—is slow.

5 Micro Validation

In the previous section, we established that our theory can reproduce the heterogeneity in the pattern of employment transitions across different types of workers. In this section, we test the theory by examining its predictions with respect to two micro phenomena. In Section 5.1, we examine the predictions of the theory with respect to the earnings losses of displaced workers—which are known to be large and very persistent. In Section 5.2, we examine the predictions of the theory with respect to the relationship between the average UE rate and unemployment duration—a relationship which is known to be sharply negative. On both accounts, we find that the predictions of the theory line up well with the data.

5.1 Earning losses of displaced workers

It is known that the earnings losses of displaced workers—i.e. workers who lose a high-tenure job—are large and persistent (see, e.g., Jacobson, Lalonde and Sullivan 1993 or Flaeen, Shapiro and Sorkin 2019) and they are even larger during recessions (see, e.g., Davis and von Wachter 2011). The ability to reproduce the magnitude and persistence of earnings losses for displaced workers is a key test for any search theory of the labor market, since such losses capture the amount of search capital embodied in a firm-worker match that survived for several years and the speed at which a worker can recoup such capital after losing it. The test is even more critical for our theory of heterogeneous workers, as the theory implies that different types follow a very different search process.

Using the LEHD over the period 1997-2008, we identify the workers who have been employed by a particular firm for a minimum of three years and who have subsequently moved from that firm to unemployment. We refer to these workers as displaced workers. For each displaced worker, we compute their pre-displacement earnings as the average of their quarterly earnings in the year prior to the displacement. Since the exact timing of the displacement event within a

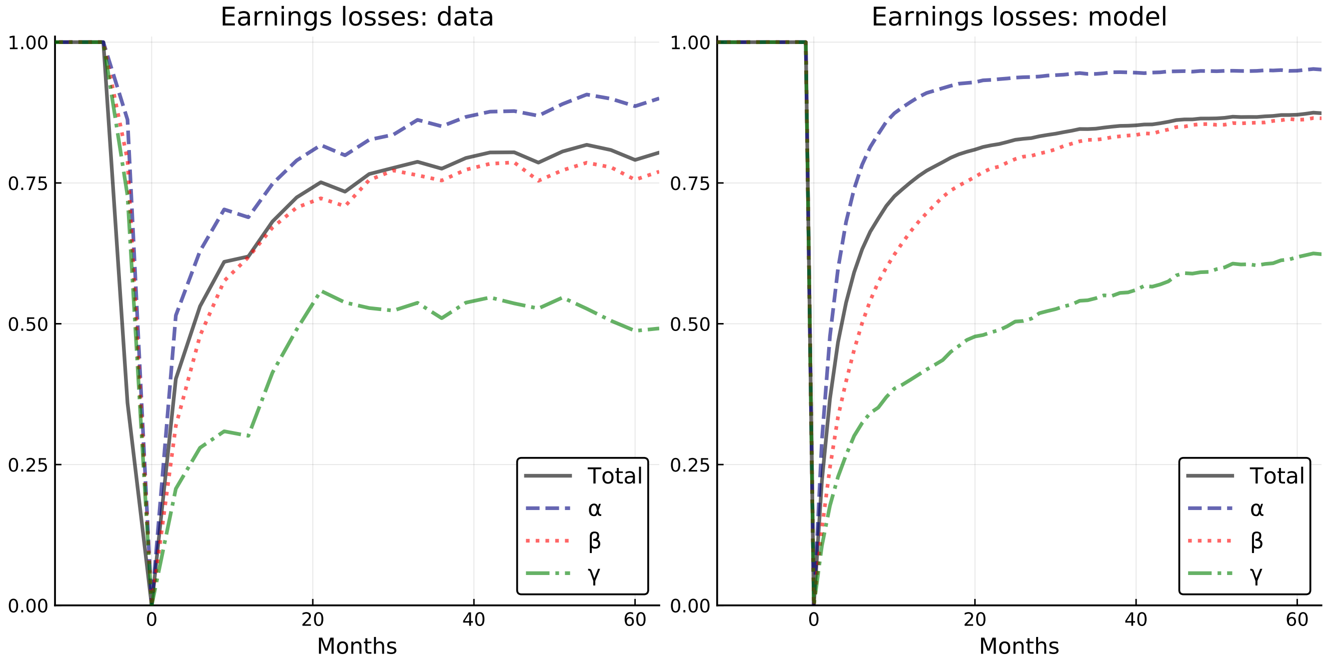

Figure 3: Earnings losses from job separation in steady state.

quarter is unknown, we impute no earnings during the displacement quarter. We then compute their post-displacement earnings for all quarters $t = 1, 2, \ldots, 20$ after the displacement episode.

In the left panel of Figure 3, we plot the ratio of the post-displacement earnings to the pre-displacement earnings averaged across all displaced workers (solid line), across α-workers (dashed line), β-workers (dotted line) and γ-workers (dash-dotted line). The average earnings losses for displaced workers are sizeable and quite persistent.[12] Six quarters after the displacement, the earnings losses are about 30%. Fourteen quarters after the displacement, the earnings losses are still about 20%. The earnings losses, however, are very different for different types. For α-workers, earnings losses are smaller and more transitory than average (20% after six quarters, and 10% after fourteen). For γ-workers, earnings losses are much larger and much more persistent than average (50% after six quarters, and still about 50% after fourteen quarters). For β-workers, earnings losses are close to the average. The right panel of Figure 3 plots the earnings losses predicted by the theory, which—as one can see—are very similar to those observed in the data.[13]

The theory provides a simple explanation for the magnitude and persistence of the earning losses for different types. Consider an α-worker who has been in the same job for more than 3 years. This worker is likely to be in a stable job, i.e. a job with quality above Q. When the worker moves into unemployment, he is likely to find a new job quickly—as α-workers have a UE rate of about 30% per month. When the worker finds a new job, he is likely to find another stable job—as α-workers have about a 50% chance of sampling a job with quality $z \geq Q$. Overall, once an α-worker is displaced from a high-tenure job, he is likely to quickly find a new

[12]In order to facilitate the comparison between data and model, we plot monthly earnings. We compute monthly earnings as a linear interpolation of quarterly earnings.

[13]Also here we approximate a worker's earnings with his productivity.

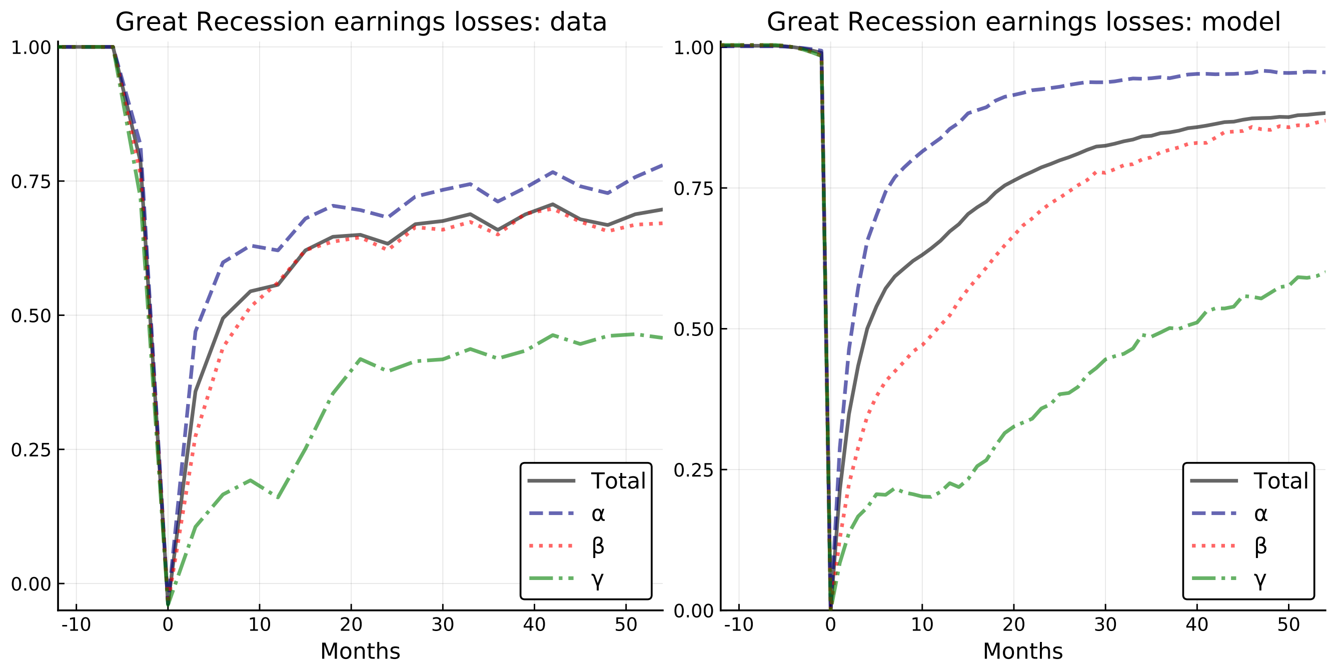

Figure 4: Earnings losses from job separation during the Great Recession.

job with a quality that is similar to the quality of his old job. Now consider a γ-worker who has been in the same job for more than 3 years. This worker is also likely to be in stable job. When the worker moves into unemployment, he needs much more time to find a new job—as γ-workers have a UE rate of about 10% per month. When the worker finds a new job, he is very unlikely to find another stable job—as γ-workers have only a 15% chance of sampling a job with quality $z \geq Q$. Most likely, after observing the quality of the new job, the worker moves back into unemployment and resumes his search for the right-tail of the distribution. Overall, once a γ-worker is displaced from a high-tenure job, it takes him a long time to find another job of the same quality as the one that he lost. Notice that, for all types of workers, the theory predicts that earnings losses will never be entirely erased, as the ergodic distribution of workers is stochastically dominated by the distribution of workers whose match has survived for more than 3 years.

In Figure 4, we construct the analogue of Figure 3 for the Great Recession. In the left panel, we plot the empirical earnings losses for displaced workers. It is easy to see that the earnings losses dissipate more slowly during the Great Recession than before, and that the impact of the recession is especially pronounced for γ-workers. In the right panel, we plot the earnings losses for displaced workers in the Great Recession that are predicted by the theory.[14] The theory correctly predicts the fact that earnings losses dissipate more slowly during the recession than before (although it fails to predict the full extent of the slow down), and that the effect is especially strong for γ-workers.

The fact that the theory does a good job at reproducing the earnings losses of different types of displaced workers is very reassuring. It suggests that the theory successfully captures the

[14]The predictions of the theory are obtained by hitting the non-stochastic steady-state of the model with the aggregate productivity shock described in Section 6.

amount of search capital embodied in matches that survive for multiple years, and the speed at which search capital is built back up after such a match is broken. It is especially noteworthy that the theory—which is calibrated only on the pattern of employment transitions—correctly predicts the stock and accumulation rate of search capital—since these objects depend directly on the distribution of match-specific productivity and wages. The success of the theory is also surprising, since reproducing the magnitude and persistence of earnings losses is a challenge for basic search-theoretic models of the labor market and typically requires adding human capital depreciation, stigmatization or other sources of scarring[15] (see, e.g., Davis and von Wachter 2011). The finding that earnings losses for displaced workers have a different magnitude and persistence for different types is interesting in its own right, as it contributes to the empirical literature on displacement.

5.2 Duration dependence of UE

It is well-known that the UE rate declines sharply with the duration of unemployment (see, e.g., Alvarez, Borovickova and Shimer 2018, Jarosch and Pillosoph 2019, Mueller, Spinnewjin and Topa 2019). The ability to reproduce the relationship between the UE rate and unemployment duration is a useful test of our theory. The UE rate at the beginning of an unemployment spell reflects the composition of workers entering unemployment. The decline of the UE rate with unemployment duration reflects the evolution of the composition of workers (i.e., dynamic selection) and, possibly, the effect of duration of duration on the UE rate of different types (i.e. true duration dependence).

Using the LEHD over the period 1997-2008, we identify workers who enter into unemployment and, for each of these workers, we record the duration of their unemployment spell. We then compute the ratio between the number of workers who have an unemployment spell that lasts $t = 1, 2, 3, 4, 5$ quarters and the number of workers who have an unemployment spell than lasts $t - 1$ quarters. The ratio gives us the average UE rate for an unemployment duration of $t - 1$ quarters.

In the top panel of Figure 5, we plot the average UE rate as a function of the unemployment duration expressed as a monthly rate (left) and the type composition of the pool of unemployment as a function of the unemployment duration (right). The monthly UE rate falls from about 22% at the beginning of an unemployment spell to about 17% after one year of unemployment, a decline of 5 percentage points. At the beginning of an unemployment spell, the pool of unemployment has 40% of α-workers, 35% of β-workers, and 25% of γ-workers. After one year of unemployment, the pool has 25% of α-workers, 35% of β-workers, and 40%

[15] Pries (2004) explains why basic search-theoretic models of the labor market have a hard time reproducing the magnitude and persistence of earnings losses for displaced workers and, relatedly, why these models have a hard time reproducing the persistence of aggregate unemployment. He suggests that both challenges may be overcome by models where firm-worker matches are assumed to be experience goods. The assumption of matches as experience goods implies heterogeneity in match quality—which increases the stock of search capital that can be accumulated—and that the search process is about both locating a match and discovering its quality—which slows down the speed at which search capital is accumulated. Not surprisingly, our model is one where matches are experience goods.

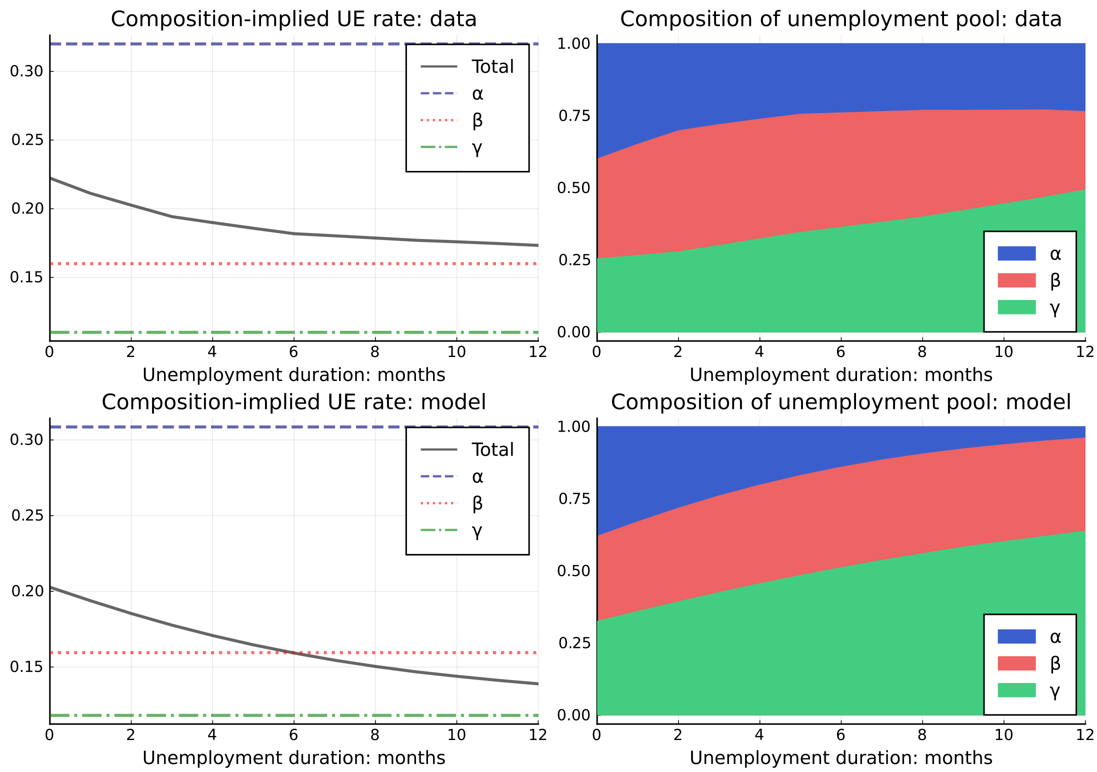

Figure 5: Left panels: type-specific and aggregate UE rates by duration; right panels: composition of unemployment pool by duration. Top panels are from the data and bottom panels are from the model.

of γ-workers.

In the bottom panel of Figure 5, we plot the predictions of the theory with respect to the average UE rate as a function of the unemployment duration (left) and the type composition of the pool of unemployment expressed as a function of the unemployment rate (right). The theory predicts that the UE rate goes from about 20% at the beginning of an unemployment spell to about 14% after one year of unemployment, a decline of about 6 percentage points. The theory predicts that, at the beginning of an unemployment spell, the pool of unemployed contains 40% of α-workers, 30% of β-workers, and 30% of γ-workers. After one year of unemployment, the theory predicts that the pool of unemployment contains 10% of α-workers, 40% of β-workers, and 50% of γ-workers. The theory correctly predicts the magnitude of the decline in the UE rate with unemployment duration, the composition of the unemployment pool at the beginning of an unemployment spell, and the direction of the change in the composition of the pool of unemployment. In light of these observations, we feel comfortable with the simplifying assumption that the type-specific UE rate is independent of unemployment duration.

The theory provides a simple explanation for the fact that the UE rate declines with the duration of an unemployment spell. According to the theory, the UE rate of each particular type of worker is independent of duration, but the UE rate of different types of workers is very different—30% for α-workers, 15% for β-workers and 10% for γ-workers. Since different types of workers have a different UE rate, the composition of the pool of unemployment shifts throughout an unemployment spell towards types with the lowest UE rate (γ) and away from types with the highest UE rate (α). In turn, the change in the composition of the unemployment pool causes the decline in the average UE rate. In other words, according to the theory, the observed decline in the average UE rate is entirely accounted for by the fact that different types of workers have a different UE rate.

The conclusion that the decline in the average UE rate is mainly due to heterogeneity in the UE rates of different workers is consistent with Mueller, Spinnewjin and Topa (2019), although we reach this conclusion through a novel route. Here, we use the entire pattern of individual transitions between employment, unemployment, and across employers in order to assign individuals to groups. We then show that heterogeneity in grouped fixed-effects can account for all of the decline in the average UE rate. Mueller, Spinnewjin and Topa (2019) use data on individual expectations about the UE rate and on the correlation between expectations and realizations to show that heterogeneity in individual UE rates accounts for nearly all of the decline in the average UE rate. It is also useful to compare our methodology with Alvarez, Borovickova and Shimer (2018). They recover individual fixed-effects in the UE rate by comparing multiple unemployment spells of the same individual, and find that heterogeneity in individual fixed-effects accounts for only a fraction of the decline in the average UE rate. In contrast, we recover fixed-effects using the overall behavior of an individual and the behavior of other workers who are similar to him.

6 Macro Measurement

In this section, we use the theory to measure the impact of aggregate productivity shocks on labor market outcomes (Section 6.1). We find that a negative aggregate productivity shock causes heterogeneous responses in the UE, EU and unemployment rates of different types of workers: small and transient responses for α-workers, and very large and very persistent responses for γ-workers. These type-specific responses aggregate up to a large and persistent increase in aggregate unemployment—mostly driven by the unemployment rate of γ-workers—and to a small and transitory decline in measured labor productivity—mostly driven by the dynamics of the composition of the employment pool. We then compare the impact of aggregate productivity shocks in our theory with the impact of the same shock in a version of the model with a representative worker (Section 6.2). We find that worker heterogeneity increases the magnitude and persistence of the response of unemployment, while it dampens the magnitude and persistence of measured productivity. Lastly, making use of the LEHD data over the period 2008-2014, we document the behavior of the type-specific unemployment rate during and after the Great Recession (Section 6.3). We find that the predictions of the theory are qualitatively very similar to what we observe in the data. We argue that type-specific productivity shocks allow us to quantitatively align the predictions of the theory to the data.

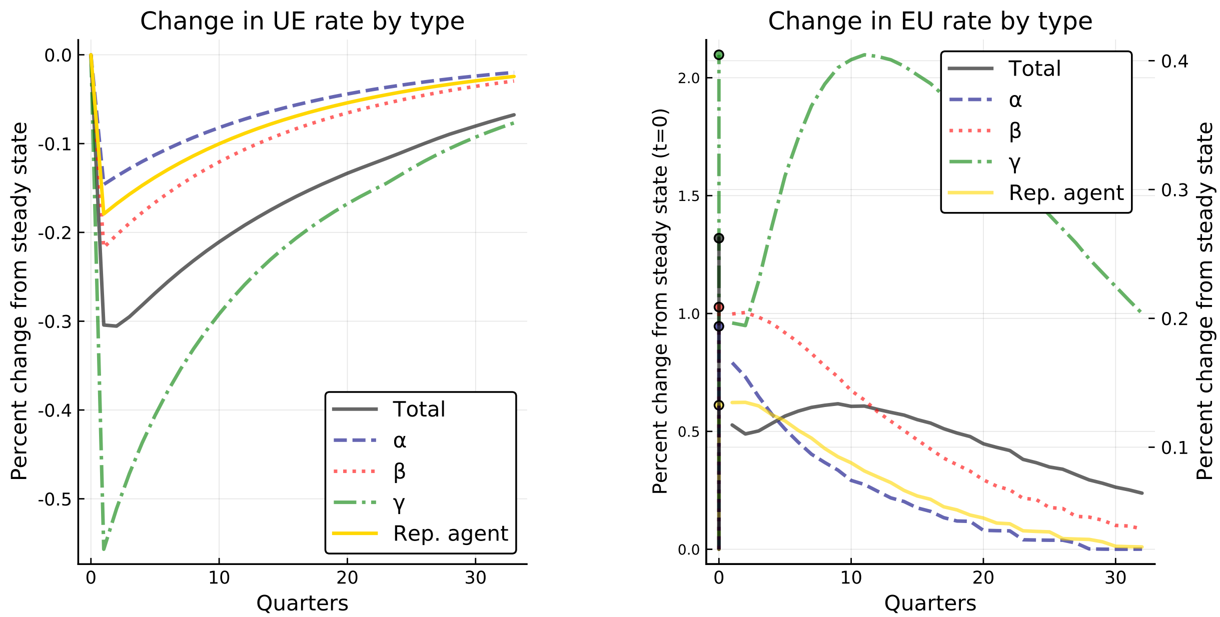

Figure 6: Left panel: model-implied UE rates by type; right panel: model-implied EU rates by type (left axis refers to the response on impact, right axis refers to everything thereafter)

6.1 Aggregate productivity shocks

We first want to use the theory in order to measure the effect of aggregate productivity shocks on the labor market. To this aim, we take the non-stochastic steady-state of the labor market and compute the response of the UE, EU and unemployment rate to a 10% negative shock to the aggregate component of productivity x with a half-life of 3 years.

Figure 6 plots the response of the UE rate (left panel) and of the EU rate (right panel) to the x-shock for different types of workers. On impact, the UE rate declines for all types of workers. However, the decline in the UE rate is very different for different types of workers. For α-workers, the UE rate declines by 15%. For β-workers, the UE rate declines by 23%. For γ-workers, the UE rate declines by 60%. The speed at which the decline in the UE rate is reabsorbed is essentially the same for all types of workers—the half-life of the response in the UE rate is about 3 years.

The logic behind the above findings is simple. The UE rate for workers of type i is proportional to the expected gains from trade for workers of type i which, in turn, are approximately proportional to the difference between expected productivity xy_i and unemployment income b_i. Hence, the same x-shock generates a larger percentage change in the gains from trade and, in turn, in the UE rate for types with a higher ratio between xy_i and $xy_i - b_i$. Given the non-linear relation between b_i in y_i, the types with the highest ratio $xy_i/(xy_i - b_i)$ are the types with the lowest baseline productivity y_i. That is, γ-workers. Conversely, the types with the lowest ratio $xy_i/(xy_i - b_i)$ are the types with the highest baseline productivity. That is, α workers.[16]

[16]This is the same logic behind the observation that the elasticity of the UE rate in the baseline search-theoretic model of Pissarides depends on the size of the gains from trade. If the gains from trade are large, as in Shimer (2005), the elasticity of the UE rate is small. If the gains from trade are small, as in Hagedorn and

Let us turn to the EU rate. On impact, the EU rate increases for all types of workers. This is because, for all types of workers, the shock to x reduces the value of employment relative to the value of unemployment and, hence, increases the reservation quality R. The magnitude and the persistence of the increase in the EU rate are, however, very different for different types of workers. Consider α-workers. On impact, the shock leads to a one-time 250% increase in the EU rate—which is caused by the destruction of existing matches with quality z that falls below the new and higher reservation quality. After the impact, the shock leads to an increase in the EU rate of about 20%–which is caused by the destruction of matches of unknown quality who are discovered to be below the new and higher reservation quality. The effect of the shock dissipates very quickly. Now, consider γ-workers. On impact, the shock leads to a one-time 500% increase in the EU rate. After the impact, the shock leads to an increase in the EU rate that starts at 20%, peaks at 40% after 10 quarters, and then dissipates very slowly.

The logic behind the above findings is relatively simple. The immediate effect of the shock on the EU rate depends on the density of the cross-sectional distribution of employed workers around the reservation quality. This density is lower for α than for γ-workers, because the flow into marginal matches—which depends on the flow of new matches relative to the stock of matches—is much lower for α than for γ-workers. The post-impact effect of the shock on the EU rate depends on the increase in the number of matches that are discovered to be below the reservation quality and, in turn, on the increase in the flow of new matches relative to the stock. The increase is smaller for α than for γ-workers because, on impact, the shock moves a smaller fraction of α than γ-workers into unemployment and, eventually, into new matches. The persistence of the effect of the shock on the EU rate depends on the search process of displaced workers. As explained before, a displaced α-worker is likely to go through only one unemployment spell before finding a stable match. In contrast, a displaced γ-worker goes through multiple spells of employment and unemployment before finding a stable match.

The left panel in Figure 7 plots the response of the unemployment rate for different types of workers. For α-workers, the unemployment rate increases by 2 percentage points and it is re-absorbed quickly (the half-life is about 1 year). For β-workers, the unemployment rate increases by 5 percentage points and it is re-absorbed more slowly (the half-life is about 3 years). For γ-workers, the unemployment rate increases by 20 percentage points and it is re-absorbed very slowly (the half-life is close to 6 years). The difference in the magnitude and the persistence of the increase in the unemployment rate for different types is a direct consequence of the difference in the magnitude and persistence of the response of the type-specific UE and EU rates.

The right panel in Figure 7 plots the response of the composition of the unemployment pool. At the non-stochastic steady state, the unemployment pool has about 22% α-workers, 27% β-workers, and 51% γ-workers. Since the shock increases disproportionately the unemployment rate of γ-workers, the composition of the unemployment pool tilts towards γ and away from α and β-workers. Since the increase in the unemployment rate decays more slowly for γ-workers, the composition of the unemployment pool remains tilted towards γ-workers for a long period

Manovskii (2009), the elasticity of the UE rate is large.

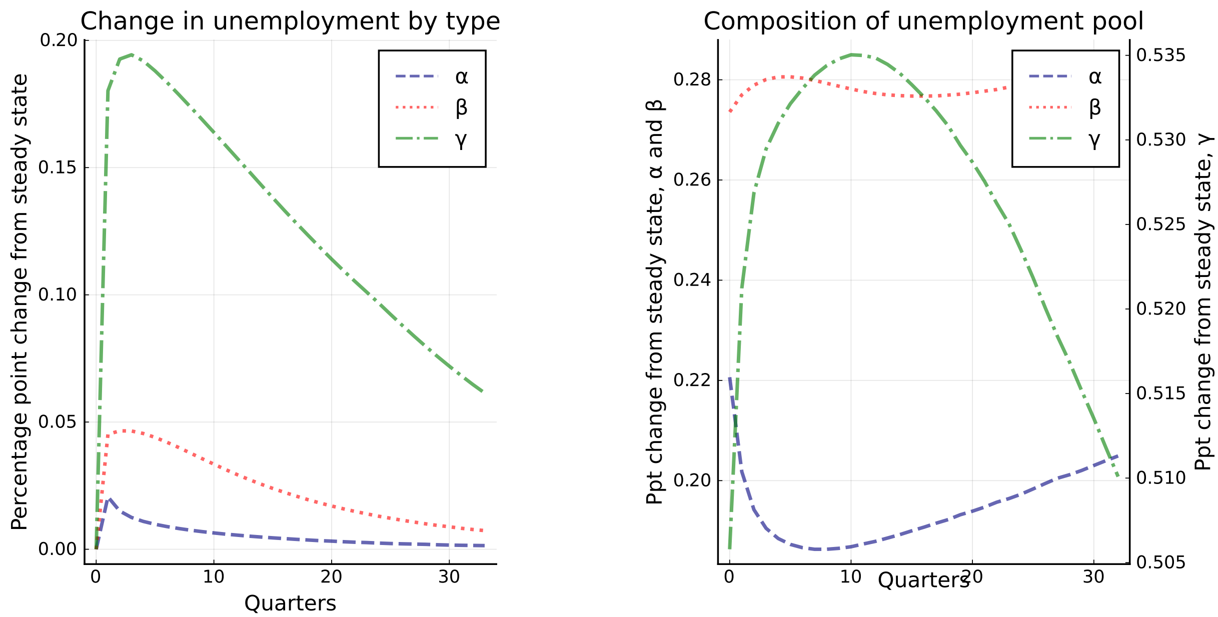

Figure 7: Model-implied unemployment rates and composition in response to a 10% decline in aggregate productivity

of time.

The changing composition of the unemployment pool helps us understand the aggregate response of the labor market to the productivity shock. The black solid line in the left panel of Figure 6 plots the response of the aggregate UE rate. On impact, the aggregate UE rate declines by about 30%. This is approximately the average of the decline in the UE rate of different types weighted by the contribution of the types to the total UE rate. Over time, the decline in the aggregate UE rate dissipates, but more slowly than any of the type-specific UE rates. Indeed, the half-life of the aggregate UE rate is about 5 years, while the half-life of the type-specific UE rate is about 3 years. This phenomenon is caused by the changing composition of the unemployment pool. As the increase in the unemployment rate for γ-workers is reabsorbed more slowly than for the other types, the weight on the response of the UE rate of γ-workers in the response of the aggregate UE rate increases. Since the response of the UE rate for γ-workers is larger than for the other types, the aggregate UE rate dissipates more slowly than any of the type-specific rates.

The black solid line in the right panel of Figure 6 plots the response of the aggregate EU rate. The increase in the aggregate EU rate is an average of the increase in the EU rate of different types. Over time, the aggregate EU rate dissipates, but more slowly than the any of the type-specific EU rates. The intuition for this phenomenon is simple. Over time, the α and β-workers that were displaced by the shock return to stable matches, while the γ-workers that were displaced by the shock experiment and leave several matches. Hence, over time, the aggregate response of the EU rate is entirely driven by γ-workers.

The aggregate behavior of the unemployment rate and of labor productivity is displayed in Figure 8. On impact, the aggregate unemployment rate increases by 5.4 percentage points,

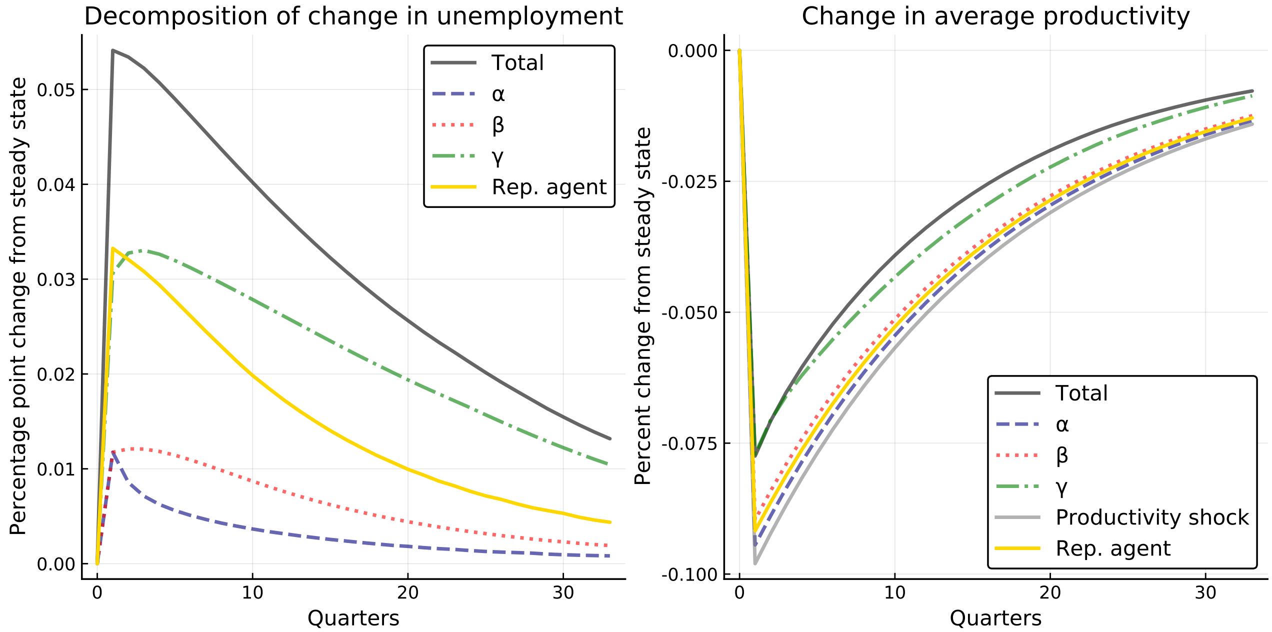

Figure 8: Left panel: decomposition of the response of the aggregate unemployment rate; right panel: aggregate and type-specific productivity of labor

with 25% of the increase due to the rise in the unemployment of α-workers, 21% due to the rise in the unemployment of β-workers, and 54% due to the rise in the unemployment of γ-workers. The increase in the unemployment of α and β workers is reabsorbed much faster than the increase in the unemployment of γ workers. Hence, the aggregate unemployment rate is eventually entirely due to the excess unemployment rate of γ-workers and, like the excess unemployment of γ-workers, it dissipates slowly. Indeed, the half-life of the increase in the aggregate unemployment rate is 5 years.

Now, turn to labor productivity (i.e. output per employed worker). On impact, labor productivity falls, but by less than the aggregate component of productivity. Indeed, the initial decline in labor productivity is 8%, while the initial decline in the aggregate component of productivity is 10%. Over time, labor productivity recovers, and it does so more quickly than both the aggregate component of productivity and the aggregate unemployment rate. Indeed, the half-life of the decline of labor productivity is 2 years, while the half-life of the aggregate productivity shock is 3 years and the half-life of the increase in the aggregate unemployment rate is 5 years.

The intuition behind the above findings is simple. The decline in labor productivity is muted because of a double cleansing effect. Within a type, the workers who are displaced by the shock are those in matches with relatively low quality. Across types, the workers who are displaced by the shock are disproportionately γ-workers, who have the lowest productivity among all types.[17] Both effects imply that the workers who survive the shock are positively selected and, hence, labor productivity declines less than the aggregate component of productivity. Over

[17] The within-type cleansing effect is also active in models with homogeneous workers and match-quality heterogeneity, such as Mortensen and Pissarides (1994) or Menzio and Shi (2011).

time, the displaced workers who first find a new stable match are α and β-workers. Hence, over time, the composition of the employment pool tilts further towards α and β-workers and away from γ-workers. Since α and β-workers are more productive than γ-workers, labor productivity recovers faster than the underlying shock.

Let us summarize our findings. First, aggregate unemployment is sensitive to labor productivity fluctuations. Indeed, the semi-elasticity of the aggregate unemployment rate with respect to labor productivity is approximately 7. Second, the recovery is jobless, in the sense that the increase in aggregate unemployment dissipates more slowly than the decline in labor productivity. Indeed, the half-life of the increase in the aggregate unemployment rate is 5 years, while the half-life of the decline in labor productivity is only 2 years. Third, the response of labor productivity is muted, in the sense that the decline in labor productivity is smaller and more transient than the shock to the aggregate component of productivity.[18]

6.2 Representative worker

In order to quantify the role played by worker heterogeneity in shaping the response of the labor market to an aggregate productivity shock, we consider a version of the model with only one type of worker. We refer to this version of the model as the Representative Worker Model (RWM), and to the baseline model as the Heterogeneous Worker Model (HWM). We calibrate RWM using the same targets used to calibrate HWM, except that all the type-specific targets are aggregated. We then use the RWM to measure the response of the labor market to a 10% negative shock to the aggregate component of productivity with a half-life of 3 years.

The yellow solid line in the left panel of Figure 6 plots the response of the UE rate to the shock in RWM. The UE rate declines by 18% on impact and recovers with a half-life of 3 years. The decline in the UE rate in RWM is smaller than in HWM because the elasticity of the UE rate with respect to the aggregate component of productivity is a convex function of $xy_i/(xy_i - b_i)$. Therefore, the response of the UE rate for the representative worker is smaller than the average response of the UE rate for heterogeneous workers. The decline in the UE rate recovers more quickly in RWM than in HWM because, with a representative worker, there is no change in the composition of the unemployment pool that slows down the recovery of the UE rate. Overall, the response of the UE rate in RWM is much closer to the response of the UE rate for α-workers than to the response of the average UE rate in HWM.

The yellow solid line in the right panel of Figure 6 plots the response of the EU rate in RWM. On impact, the EU rate increases by about 50%. After the impact, the EU rate increases by 12% and recovers with a half-life of 3 years. Again, the response of the EU rate in RWM is closer to the response of the EU rate for α-workers than to the response of the average EU rate in

[18] The magnitude of the response of the UE, EU and unemployment rates relative to the magnitude of the labor productivity decline are in the same ballpark as what we observe in typical US recessions. The relative volatility of the UE rate is approximately 3, the relative volatility of the EU rate is approximately 4, and the relative volatility of the unemployment rate is approximately 7. Using BLS data, the relative volatility of the UE and EU rates are about 6, and the relative volatility of the unemployment rate is about 9 (see, e.g., Menzio and Shi 2011, Table 1). Keep in mind, however, our notion of unemployment and, hence, of UE and EU rates is not the same as in the BLS.

HWM. This is because the representative worker—who is calibrated to match the distribution of job durations in the aggregate—samples from a match-quality distribution that looks a lot like the one for an α-worker.

The responses of unemployment and labor productivity in RWM are plotted in yellow in the left and right panels of Figure 8. The shock causes an increase in the unemployment rate of about 3 percentage points that dissipates with a half-life of 3 years. That is, the increase in the unemployment rate is smaller and less persistent in RWM than HWM. These differences are an immediate consequence of the difference in the response of the UE and EU rates in RWM and HWM. The shock causes a decline in labor productivity of about 9% that dissipates with a half-life of 3 years. That is, the decline in labor productivity is larger and more persistent in RWM than HWM. Without worker heterogeneity, the negative productivity shock has only a cleansing effect only on the quality of surviving matches and not on the quality of the surviving workers. Hence, the decline in labor productivity is larger in RWM than HWM. Moreover, without worker heterogeneity, the composition of the employment pool does not change over time. Hence, the half-life of the decline in labor productivity is longer in RWM than HWM.

Let us summarize our findings. First, worker heterogeneity increases the magnitude and persistence of the response of UE, EU and unemployment rates to an aggregate productivity shock. Specifically, worker heterogeneity almost doubles the magnitude and half-life of the response of the UE and EU rates, and it increases the magnitude of the response of the unemployment rate by 2 percentage points and its half-life by 2 years. Second, worker heterogeneity reduces the magnitude and persistence of the response of labor productivity to an aggregate productivity shock. Specifically, worker heterogeneity reduces the magnitude of the response of labor productivity by 15% and lowers its half-life by 1 year. Therefore, worker heterogeneity contributes to the elasticity of unemployment to labor productivity fluctuations, to the joblessness of recoveries, and to the dampening of labor productivity fluctuations.

Here is an impressionistic description of our analysis: The aggregate response of the labor market in RWM is similar to the response that would obtain if the economy was populated only by α-workers—who are the majority of workers in the economy. In contrast, the aggregate response of the labor market in HWM is similar—especially during the later stages of the recovery—to the response that would obtain if the economy was populated mainly by γ-workers—who are a small, non-representative fraction of workers in the economy. Since α and γ workers are very different, the response to the same shock is very different in RWM and HWM. In particular, since α-workers have larger gains from trade than γ-workers, the magnitude of the response of the labor market is smaller in RWM than in HWM. Since the search process of α-workers is quick and the search process for γ-workers is slow, the labor market recovers faster in RWM than in HWM.

6.3 The Great Recession

Even though actual recessions are likely caused by a multiplicity of shocks (e.g., productivity shocks, demand shocks, financial shocks, etc…), we believe that it is still useful to compare the

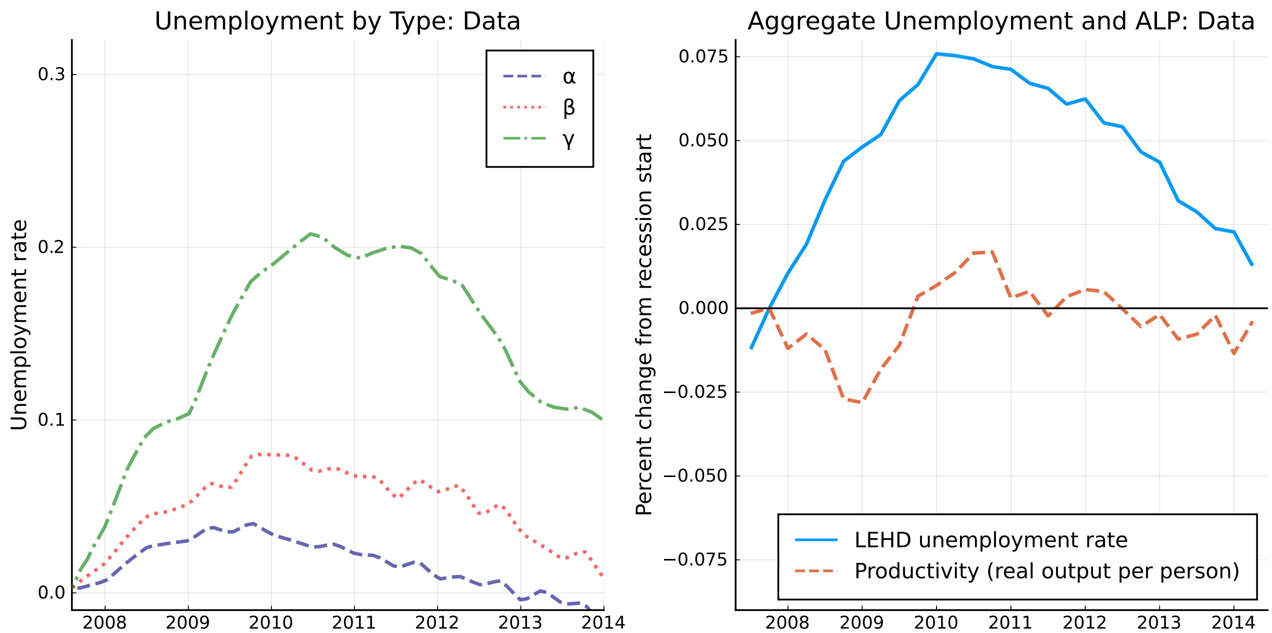

Figure 9: Left panel: increase in unemployment rates for each worker type during the Great Recession in the data; right panel: LEHD unemployment and average labor productivity during the Great Recession in the data

theory's predictions about the response of the labor market to a productivity shock with the actual behavior of the labor market during a recession. To this aim, we make use of the fact that our extract of the LEHD covers the period 1997-2014, which includes the Great Recession of 2008-2009 and its aftermath. Using the LEHD data from 2008 to 2014, we can construct time-series for the aggregate unemployment rate, and for the unemployment rate of different types.[19]

The left panel of Figure 9 plots the unemployment rate of different types of workers during and after the Great Recession. More specifically, the panel plots the unemployment rate of different types of workers during the period 2008-2014 net of their unemployment rate in the last quarter of 2007. The unemployment rate of α-workers increases by 3 percentage points over the period 2008-2010, and quickly falls down afterwards. By the end of 2012, the unemployment rate of α-workers is back to its pre-recession level. The unemployment rate of β-workers increases by 8 percentage points over the period 2008-2010, and falls down afterwards. The unemployment rate returns to its pre-recession level by 2014. The unemployment rate of γ-workers increases by 21 percentage points between 2008 and 2010 and, afterwards, it falls slowly back. By 2014, the unemployment rate of γ-workers is still 10 percentage points higher than before the recession.

[19]In order to construct the type-specific unemployment rate during and after the Great Recession, we need to make some adjustments so as to take care of attrition and time-trends in labor market participation. First, we take workers who are in our sample in a quarter T prior to 2018 and track their unemployment rate in quarter $T+t$, $t = 0, 1, 2, \ldots$ Second, to control for trends in labor market participation, we estimate a linear time trend in the unemployment rate of a cohort after t quarters. The linear time trend is allowed to vary by type. Lastly, we measure the excess unemployment rate of the cohort of workers who are in our sample in the last quarter of 2017 as their unemployment rate net of the unemployment rate forecasted using the behavior of prior cohorts and the estimated linear trends.

The right panel of Figure 9 plots the behavior of the aggregate excess unemployment and of labor productivity—measured as the percentage deviation of output per worker from trend. The aggregate unemployment rate increases by 7.5 percentage points from 2008 to 2010, and then slowly falls back towards its pre-recession level. Labor productivity falls by 2.5 percent from trend from 2008 to 2009 and then recovers quickly, returning to trend by 2010.

The picture in the left panel of Figure 9 is dramatic. The increase in the unemployment rate of γ-workers is 3 times as large as the increase in the unemployment rate of β-workers, and about 7 times as large as the increase in the unemployment rate of α-workers. Moreover, excess unemployment would have been essentially all reabsorbed by 2014 had it not been for γ-workers—who still had to reabsorb half of their excess unemployment.

The picture in the left panel of Figure 9 is qualitatively similar to the one in the left panel of Figure 7—which plots the response of the type-specific unemployment rate to a 10% negative productivity shock. First, in response to the productivity shock, the unemployment rate increases by 20 percentage points for γ-workers, by 5 percentage points for β-workers, and by 2.5 percentage points for α-workers. That is, the theory predicts that the productivity shock generates responses in the unemployment rate of different types that are of the same order of magnitude as what we observe in the Great Recession. Second, the theory predicts that the productivity shock generates responses in the unemployment rate that have different persistence for different types—least persistent for α-workers and most persistent for γ-workers. This is precisely what we observe in the aftermath of the Great Recession. Lastly, the theory predicts that the productivity shocks generates a decline in labor productivity that is much less persistent than the increase in unemployment. This is also consistent with what we see in the data.

We believe that the qualitative fit between the predictions of the theory in response to a negative productivity shock and the behavior of the US labor market during and after the Great Recession provides a powerful validation of our framework. The fit shows that the theory correctly identifies the relative susceptibility of different types of workers to negative aggregate shocks, and it correctly predicts the speed of the process through which displaced workers of different types return to stable employment. The theory, however, appears to mis-specify the precise nature of the shock or to omit some additional shocks—since it predicts a decline in labor productivity of 7.5%, while in the data labor productivity fell by only 2.5%.

In the context of our theory, a natural alternative to a shock to the aggregate component of productivity x are shocks to the type-specific components of productivity y_i. One interpretation of y_i-shocks is that different types of workers are employed in different tasks or different roles and that these differences make their productivity more or less sensitive to technology shocks. Another interpretation of y_i-shocks is that other shocks, say financial or demand shocks, impact the value of their output differently. Leaving aside issues of interpretation, here we simply want to understand whether we can find a series of y_i-shocks that is able to reproduce the magnitude and persistence of the increase in the type-specific and aggregate unemployment rate and the magnitude and persistence of the decline in labor productivity observed during the Great Recession.

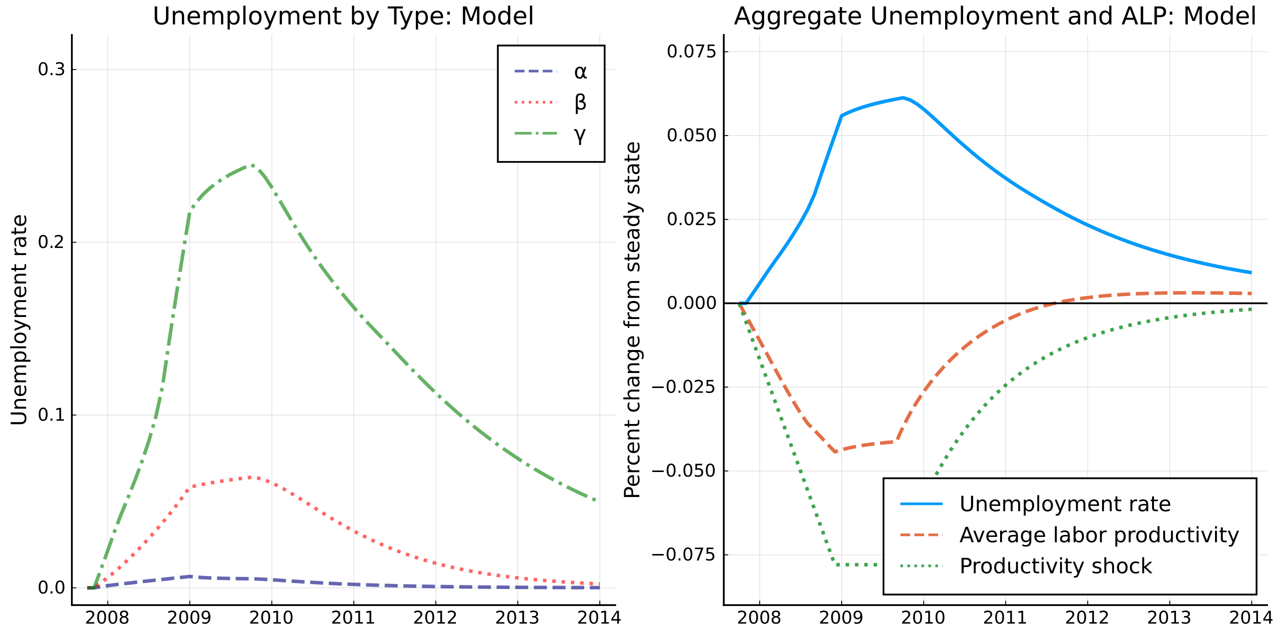

Figure 10: Left panel: increase in unemployment rates for each worker type during the Great Recession in the model; right panel: LEHD unemployment and average labor productivity during the Great Recession in the model

We feed into the model a series of shocks to the type-specific components of productivity that are perfectly correlated across types, but have different magnitudes. We assume that the type-specific components of productivity fall linearly from the beginning of the fourth quarter of 2007 to the end of 2008, they remain constant throughout 2009, and then they start recovering back to their normal levels at an exponential rate of 10% per month. The shocks to the type-specific component of productivity are smaller for α-workers than for β and γ-workers. We assume that, at the through, the the type-specific productivity is 5% lower than normal for α-workers and 12% lower than normal for β and γ-workers.

The left panel of Figure 10 plots the response of the unemployment rate for different types of workers. Between 2008 and 2010, the unemployment rate increases by 1 percentage points for α-workers,[20] by 6.5 percentage points for β-workers, and by 23 percentage points for γ-workers. After that, the increase in the unemployment rate is reabsorbed for all types of workers, but at very different speeds. The increase in the unemployment rate of α-workers is fully reabsorbed in 2012. The increase in the unemployment rate of β-workers is fully reabsorbed in 2013. For γ-workers, the increase in the unemployment rate is reabsorbed so slowly that it is still 7 percentage point above its pre-recession level in 2014.

The right panel plots the response of aggregate unemployment and labor productivity. Aggregate unemployment increases by 5.5 percentage points between 2008 and 2010. From 2010 onwards, aggregate unemployment falls slowly back towards its pre-recession level. At the beginning of 2014, aggregate unemployment is still 1 percentage point higher than before the

[20]The increase in the unemployment rate of α-workers is smaller in the model than in the data. However, the difference is small—in percentage point terms—and may very well be due to errors in the empirical measurement of the unemployment rate of α-workers.

recession. Both the magnitude and the persistence of the increase in aggregate unemployment are driven by the response of the unemployment rate for γ-workers. The response of labor productivity is not the mirror image of aggregate unemployment. Between 2008 and 2009, labor productivity declines by 4%. Afterwards, labor productivity recovers very quickly and it is back to its pre-recession level midway through 2011. The decline in labor productivity is so small because of a strong cleansing effect across types. Indeed, since the decline in productivity is larger for β and γ-workers, the composition of the employment pool tilts strongly towards α-workers who are the most productive. The decline in labor productivity is so transitory because, during the recovery, the composition of the employment pool tilts even more towards α-workers.

A comparison between Figures 9 and 10 reveals that shocks to the type-specific component of productivity that are perfectly correlated across types but larger for less productive types allow the model to reproduce quite well the behavior of the US labor market during and after the Great Recession.[21] These type-specific productivity shocks increase the magnitude and persistence of the response of unemployment and they lower the magnitude and persistence of the response of labor productivity. They do so by leveraging one of the theory's main insight: the disconnect between the types that drive the dynamics of aggregate unemployment (γ-workers) and the types that drive the dynamics of labor productivity (α-workers).

7 Conclusions

Most research in macro-labor is carried out using search-theoretic models of the labor market where the employment transitions of all workers are realizations of the same stochastic process (e.g., Mortensen and Pissarides 1994, Shimer 2005, Hall 2005, Menzio and Shi 2011). In line with the representative worker assumption, these models are estimated using either aggregate data (e.g., unemployment, and aggregate flows between employment and unemployment) or cross-sectional cuts of micro data (e.g., distribution of unemployment spell durations, distribution of job durations, etc…). With this paper we hope to have shown—by means of a prototype—that it is possible and fruitful for macro-labor to move beyond from the representative worker assumption. In terms of available data, there are now several panel datasets that record the employment histories of millions of workers for many consecutive years (e.g., LEHD in the US, IDA in Denmark, etc…). In terms of estimation methods, the Grouped Fixed-Effects method of Bonhomme, Lamadon and Manresa (2021) now provides a way to summarize workers' heterogeneity by grouping workers into types and then by estimating the type-specific parameters of a model with heterogeneous types—an approach that is computationally much lighter than maximum likelihood. In terms of models, the directed search framework of Menzio and Shi (2011) can handle rich heterogeneity because its equilibrium has the property of being

[21] We only show that there exist type-specific productivity shocks that allow the theory to reproduce the observed behavior of the labor market during the Great Recession. Since we do not have direct measures of type-specific productivity shocks, we cannot say whether these shocks are the actual force behind the Great Recession. That is, we are doing a fitting exercise rather than a macro measurement in the spirit of Kydland and Prescott (1983).

block recursive. Substantively, in this paper, we showed that workers' heterogeneity matters to understand macro phenomena like the business cycle. In doing so, we vindicated and quantified some of the conjectures from Pries (2008).

We believe that the approach outlined in this paper can open the door for much exiting research. In terms of data, it would be interesting to use panel datasets that contain even more detailed information about workers' histories, such as the exact start and end dates of each job. In terms of estimation, it would be interesting to attempt to classify workers along more dimensions than only employment transitions, such as earnings, occupation, etc... Also, it would be interesting to use longer panel datasets, in an attempt to understand whether a worker's type is a trait that is permanent or whether it does evolve over time. In terms of substantive questions, it is natural to wonder about the welfare consequences of aggregate shocks—and, in turn, the cost and benefits of monetary, fiscal and labor market policies—once workers' heterogeneity is taken into account. It would also be interesting to see whether the distribution and nature of worker types changes over the years and whether these changes may be partly responsible for the decline in labor market dynamism and for changes in the nature of the business cycle.[22]

[22]In a companion paper (Gregory, Menzio and Wiczer 2020), we use a version of the model presented in this paper to forecast the dynamics of non-employment after the pandemic recession of 2020.

References

[1] Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in *Producer Dynamics: New Evidence from Micro Data*, National Bureau of Economic Research, Inc, NBER Chapters, 149–230.

[2] Ahn, H., and J. Hamilton. 2019. "Heterogeneity and Unemployment Dynamics" *Journal of Business & Economic Statistics.*

[3] Alvarez, F., K. Borovickova, and R. Shimer. 2018. "Decomposing duration dependence in a stopping time model." *National Bureau of Economic Research.*

[4] Bagger, J., F. Fontaine, F. Postel-Vinay, and J. Robin. 2014. "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics." *American Economic Review*, 104: 1551-1596.

[5] Bonhomme, S., T. Lamadon, and E. Manresa. 2021. "Discretizing Unobserved Heterogeneity," *Econometrica.*

[6] Bonhomme, S., T. Lamadon, and E. Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3): 699-739.

[7] Cahuc, P. F. Postel-Vinay and J. Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence." *Econometrica*, 74: 323-364.

[8] Davis, S., and T. von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity.*

[9] Diamond, P. 1982. "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 49: 217-227.

[10] Flaeen, A., M. Shapiro, and I. Sorkin. 2019. "Reconsidering the Consequences of Worker Displacements: Firm versus Worker Perspective." *AEJ: Macroeconomics*, 11 (2): 193-227.

[11] Fujita M., and G. Ramey. 2007. "Job matching and propogation." *Journal of Economic Dynamics and Control*, 31(11): 3671-3698.

[12] Fujita M., and G. Moscarini. 2017. "Recall and Unemployment." *American Economic Review*, 107(12): 3875-3916.

[13] Gali, J., and T. van Rens. 2019. "The Vanishing Procyclicality of Labor Productivity," Manuscript, University of Warwick.

[14] Gertler, M., and A. Trigari. 2009. "Unemployment Fluctuations with Staggered Nash Wage Bargaining." *Journal of Political Economy*, 117: 38-85.

[15] Golosov, M., and G. Menzio. 2020. “Agency Business Cycles.” *Theoretical Economics*, 15:123-158.

[16] Gregory, V., Menzio, G., and D. Wiczer. 2020. “Pandemic Recession: L- or V-Shaped?” *Quarterly Review*, 40: 1–28.

[17] Hagedorn, M., and I. Manovskii. 2008. “The Cyclical Behavior of Unemployment and Vacancies Revisited.” *American Economic Review,* 98: 1692–706.

[18] Hall, R. 2005. “Employment Fluctuations with Equilibrium Wage Stickiness,” *American Economic Review*, 95: 50-65.

[19] Hall, R. 2017. “High Discounts and High Unemployment,” *American Economic Review*, 107: 305-330.

[20] Hall, R., and M. Kudlyak. 2019. “Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics Paper.” FRB San Francisco WP No. 19-05.

[21] Hall, R., and P. Milgrom. 2008. “The Limited Influence of Unemployment of the Wage Bargain.” *American Economic Review,* 98: 1653–74.

[22] Jacobson, L., R. LaLonde, and D. Sullivan. 1993. “Earnings Losses of Displaced Workers.” *American Economic Review*, 83: 685-709.

[23] Jarosch, G., and L. Pillosoph. 2019. “Statistical Discrimination and Duration Dependence in the Job Finding Rate.” *Review of Economic Studies*, 86: 1631–1665.

[24] Kaplan, G., and G. Menzio. 2016. “Shopping Externalities and Self Fulfilling Unemployment Fluctuations.” *Journal of Political Economy*, 124: 771-825.

[25] Karahan, F., S. Ozkan, and J. Song. 2019. “Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital.” *FRB of New York Staff Report*.

[26] Kehoe, P., V. Midrigan and E. Pastorino. 2019. “Debt Constraints and Employment.” *Journal of Political Economy*, 127: 1926-1991.

[27] Kennan, J. 2010. “Private Information, Wage Bargaining and Unemployment Fluctuations.” *Review of Economic Studies*, 77: 633-664.

[28] Kydland, F., and E. Prescott. 1982. “Time to Build and Aggegate Fluctuations.” *Econometrica*, 50: 1345-1370.

[29] Ljungqvist, L., and T. Sargent. 2017. “The Fundamental Surplus.” *American Economic Review*, 107: 2630-65.

[30] Martellini, P., G. Menzio, and L. Visschers. 2020. “Revisiting the Hypothesis of High Discounts and High Unemployment.” *Economic Journal*.

[31] Menzio, G., and S. Shi. 2010. "Block Recursive Equilibria for Stochastic Models of Search on the Job." *Journal of Economic Theory*, 145: 1453-1494.

[32] Menzio, G., and S. Shi. 2011. "Efficient Search on the Job and the Business Cycle." *Journal of Political Economy*, 119: 468-510.

[33] Menzio, G., I. Telyukova, and L. Visschers. 2016. "Directed Search over the Life Cycle." *Review of Economic Dynamics*, 19: 38-62.

[34] Morchio, I. 2020. "Work Histories and Lifetime Unemployment." *International Economic Review* 61.

[35] Mortensen, D. 1970. "Job Search, the Duration of Unemployment, and the Phillips Curve." *American Economic Review*, 60(5): 847–62.

[36] Mortensen, D. and C. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies*, 61: 397-415.

[37] Mueller, A., J. Spinnewijn, and G. Topa. 2019. "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias." *American Economic Review*.

[38] Pissarides, C. 1984. "Efficient Job Rejection." *Economic Journal*, 94: 97-108.

[39] Pissarides, C. 1985. "Short-run equilibrium dynamics of unemployment, vacancies, and real wages." *American Economic Review*, 75: 676-690.

[40] Pissarides, C. 2000. *Equilibrium Unemployment Theory*. MIT Press, Cambridge, MA.

[41] Pries, M. 2004. "Persistence of employment fluctuations: A model of recurring job loss." *Review of Economic Studies*, 71: 193-215.

[42] Pries, M. 2008. "Worker heterogeneity and labor market volatility in matching models." *Review of Economic Dynamics*, 11: 664-678.

[43] Pries, M., and R. Rogerson. 2005. "Hiring policies, labor market institutions, and labor market flows." *Journal of Political Economy*, 113: 811-839.

[44] Postel Vinay, F., and J. Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica*, 70: 2295-2350.

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

[46] Wang, J. 2010. "Consistent selection of the number of clusters via crossvalidation," *Biometrika*, 97: 893–904.