ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

The Alpha Beta Gamma of the Labor Market

Authors	Victoria Gregory, Guido Menzio, and David Wiczer
Working Paper Number	2021-003E
Revision Date	October 2024
Citable Link	https://doi.org/10.20955/wp.2021.003
Suggested Citation	Gregory, V., Menzio, G., Wiczer, D., 2024; The Alpha Beta Gamma of the Labor Market, Federal Reserve Bank of St. Louis Working Paper 2021-003. URL https://doi.org/10.20955/wp.2021.003

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

The Alpha Beta Gamma of the Labor Market*

Victoria Gregory
FRB Saint Louis

Guido Menzio
New York University and NBER

David Wiczer
FRB Atlanta

August 2024

Abstract

We access a long panel dataset of US workers to document the extent to which individuals are heterogeneous with respect to their pattern of transitions across employment states. We find that heterogeneity is well approximated by three latent types: αs, βs and γs. Workers of type α leave unemployment quickly and, once they find a job, they are likely to keep it for more than 2 years. Workers of type γ find employment slowly and, once they do find a job, they are likely to leave it within 1 year. We use our empirical findings to calibrate a search-theoretic model in which workers are heterogeneous with respect to the parameters governing their employment transitions. We find that αs move quickly out of unemployment to employment because they have large gains from trade, and they are likely to stay on a job for more than 2 years because their productivity is similar in different jobs. In contrast, γs exit unemployment slowly because their gains from trade are small, and they are likely to leave a job within 1 year because they are much more productive in a small fraction of jobs than in the majority of jobs. We find that a negative shock to aggregate productivity leads to a large and persistent increase in unemployment that is mainly driven by γ-workers. The predictions of the model align well with the unemployment dynamics observed during the Great Recession.

JEL Codes: E24, O40, R11.

Keywords: Search frictions, Unemployment, Business Cycles.

*We are especially grateful to Fabian Lange, Thibaut Lamadon and Elena Manresa for their help in shaping the analysis of workers' heterogeneity, and to Jim Hamilton, Mike Pries and Richard Rogerson for providing insightful discussions of our paper. We thank Bob Hall and Mariana Kudlyak, with whom we had numerous conversations about our related projects. We are also grateful to Steve Davis, Jean Flemming, Nezih Guner, Nir Jaimovich, Loukas Karabarbounis, Simon Mongey, Iacopo Morchio, Giuseppe Moscarini, Andi Mueller, Ayşegül Şahin, Saverio Spinella, Antonella Trigari, Gianluca Violante and audience members at many conferences and seminars. The views expressed here are solely those of the authors and do not necessarily reflect those of the Federal Reserve Banks of St. Louis, Atlanta, 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.

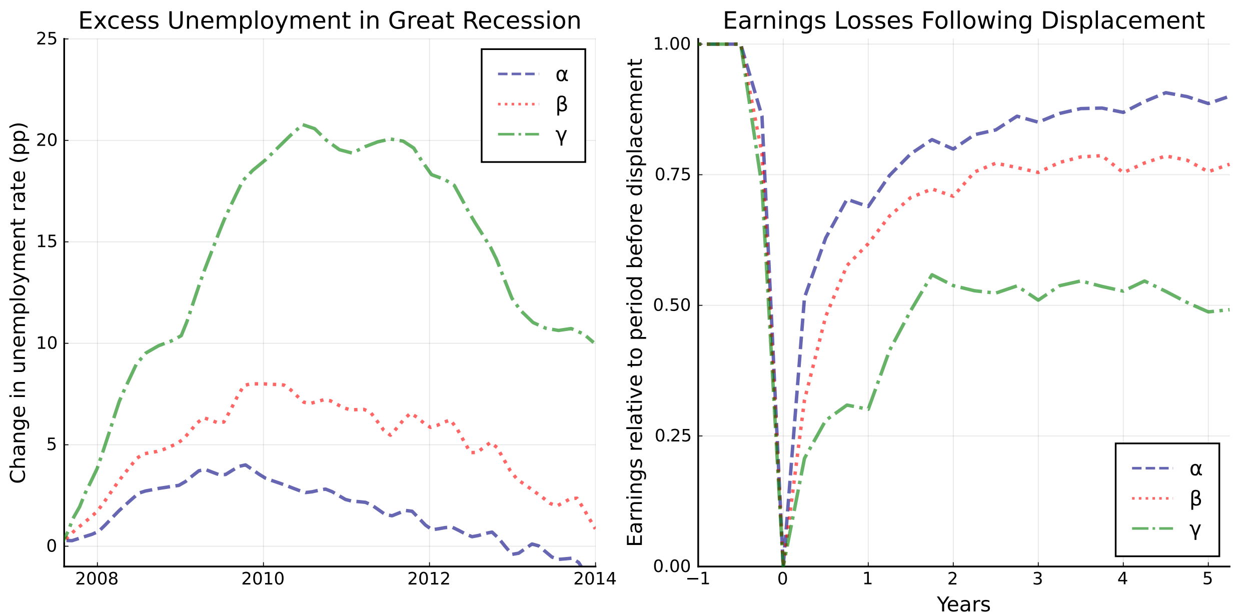

Figure 1: Left panel: excess unemployment by worker type in the Great Recession; right panel: earnings losses for high tenure workers following a job displacement.

1 Introduction

Why does unemployment increase so much in response to relatively small declines in labor productivity? And why does the increase in unemployment dissipate so slowly given that labor productivity recovers relatively quickly? In this paper, we argue that the proximate cause of these phenomena is a type of worker, which we label γ, whose unemployment increases dramatically during recessions and dissipates slowly during recoveries. This is shown in the left panel of Figure 1, which plots the excess unemployment for γ-workers during and after the Great Recession, alongside the excess unemployment for two other types of workers, αs and βs. We argue that the deeper cause of these phenomena is that the job-search process of γ-workers is qualitatively different than the job-search process of the average worker in the economy. Specifically, γ-workers move slowly from unemployment to employment, and, once they find a job, they have a low probability of keeping the job for more than 2 years. This is illustrated in the right panel of Figure 1, which plots the earnings losses for γ-workers that are displaced from a high-tenure job, alongside the earnings losses for αs and βs. Lastly, we argue that, when the economy is hit by a negative productivity shock, many γ-workers become unemployed because, due to the nature of their search process, they failed to climb the job ladder and were left in marginal jobs. Once these γ-workers become unemployed, it takes them a long time to find stable employment, as they need to experiment with several jobs and go through several spells of unemployment before settling down.

We derive our findings in three steps. First, we access a long and large panel dataset of US workers to measure the extent of workers' heterogeneity with respect to their pattern of transitions across employment states. Second, we calibrate a search-theoretic model of the labor market to match the extent of workers' heterogeneity measured in the data. Third, we use the calibrated model to measure the impact of aggregate productivity shocks on the labor market outcomes of different workers and, ultimately, to interpret the empirical behavior of the US labor market during and after the Great Recession.

There are a couple of technical hurdles involved in carrying out the three steps of the analysis. First, we need to find a parsimonious way to measure workers' heterogeneity, since estimating individual-specific parameters for about half a million workers via maximum likelihood would be too cumbersome. Second, we need to find a way to solve for the aggregate dynamics of an equilibrium model in which workers are ex-ante and ex-post heterogeneous. We tackle these hurdles by using the 2-stage Grouped Fixed Effects (GFE) method of Bonhomme Lamadon and Manresa (2021) to estimate a version of the directed search model of Menzio and Shi (2011). In the first step of the estimation, we discretize workers' heterogeneity by assigning individuals to a small number of types. In the second stage, we calibrate the model to match type-specific moments computed in the first stage. As shown in Bonhomme, Lamadon and Manresa (2021), the GFE method provides consistent estimates of individual-specific parameters as the number of individuals, the number of periods of observation and the number of types grow large. As shown in Menzio and Shi (2011), the assumption of directed search guarantees that the equilibrium of the model is block recursive—in the sense that value and policy functions do not depend on workers' heterogeneity—and hence the equilibrium of the heterogeneous-worker model can be computed as easily as the equilibrium of a model in which all workers are homogeneous.

In the first part of the paper, we document and discretize 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, 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 records. 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 βs is between the pattern for αs and γs. In terms of earnings, αs make more than βs, and about twice as much as γs. Interestingly, we find that a worker's type cannot be forecast by demographic characteristics and industry—a finding that suggests that grouping workers by demographics and industry is not the best way to approximate this form of heterogeneity.

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 current employment position. The quality of a firm-worker match is random and it is observed only after the match is formed. 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, αs 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, γs 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 γs than for αs. 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 γs and away from αs. 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. For αs, the shock leads to a small decline in the UE rate, and to a small and short-lived increase in the EU rate. As a result, the increase in the unemployment rate of αs is small and transitory. For γs, the shock leads to a large decline in the UE rate, and to a large and persistent increase in the EU rate. As a result, the increase in the unemployment rate of γs is large and persistent. Intuitively, the UE rate of γs is more sensitive to the shock because the gap between the market productivity and the value of non-market activities is smaller for γs and, hence, the shock reduces their gains from trade by a larger percentage. The EU rate of γs is more sensitive to the shock because γs are more likely to be in marginal matches. The EU rate of γs is affected for a longer time by the shock because displaced γs 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 larger and more persistent and to a decline in labor productivity that is smaller and more transitory than the underlying productivity 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 3b3s, 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.

The aggregate predictions of the model are qualitatively consistent with the data. 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 αs, by 8 percentage points for βs, and by a striking 20 percentage points for γs. Moreover, we find that the increase in the unemployment rate was reabsorbed quickly for αs (2013), less so for βs (2014), and much more slowly for γs —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 smaller and 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 γs than for αs can go a long way in realigning the quantitative predictions of the model with the actual behavior of unemployment and labor productivity.

To the best of our knowledge, our paper is the first to develop a search-theoretic model of the labor market with heterogeneous workers that is calibrated using empirical evidence on the pattern of employment transitions of individual workers, validated using a variety of micro data, and used for business cycle analysis. The main contribution of the paper is neither one particular piece of evidence or a theoretical twist, even though we decided to lead with our findings on unemployment fluctuations. The main contribution of the paper is to provide a coherent, calibrated, and tractable framework that explicitly takes into account the fact that workers are very different with respect to the speed at which they move from unemployment to employment, the frequency at which they become unemployed, the amount of time they spend on different jobs, and that this heterogeneity runs deeper than the differences in average transition rates between workers with different observable characteristics.

Different parts of the paper contribute directly to different strands of the literature. First, the documentation and discretization of workers' heterogeneity is related to the literature that tries to measure workers' fixed effects in employment transitions (e.g., Ahn and Hamilton 2020, Morchio 2019, Hall and Kudlyak 2019, Karahan, Ozkan and Song 2019, Shibata 2015). We discuss our contribution to this literature in Section 2. Second, the validation of our theoretical model of workers' heterogeneity is related to the literature studying the earnings losses of displaced workers (e.g., Jacobson, Lalonde and Sullivan 1993, Davis and von Wachter 2011, Flaeen, Shapiro and Sorkin 2019), and to the one studying the relationship between the UE rate and unemployment duration (e.g., Honore 1993, Alvarez, Borovickova and Shimer 2018, Mueller, Spinnewjin and Topa 2019). We discuss our contribution to these literatures in Section 5. Third, the empirical findings on the behavior of different types of workers during the Great

Recession offers new insights on the causes of unemployment volatility and persistence. And the finding that our model can explain quite well these findings as a response to an aggregate shock to labor productivity offers new insights on why unemployment is so volatile and persistent. Hence, the paper contributes to the literature trying to understand the cyclical behavior of the labor market through the lens of search theory (e.g., Pissarides 1985, Mortensen and Pissarides 1994, Shimer 2005, etc…). We discuss our contribution to this literature in Section 6.

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 the similarity of 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.[1] As established in Bonhomme, Lamadon and Manresa (2021), the two-step GFE method provides consistent estimates of individual-specific parameters.[2] 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. In Section 2.3, we explore the relationship between the demographic characteristics of a worker and his type. In Appendix A, we study the finite sample properties of the algorithm used to group workers into types, and show that it performs quite well given the size of our data.

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

[1] In this paper, we are interested in measuring workers' heterogeneity with respect to employment transitions and measuring the impact of this type of heterogeneity on aggregate macroeconomic fluctuations. For this reason, when we discretize workers' heterogeneity, we do not include data on earnings or demographics. Had we included such information, we would have likely ended up grouping together some workers with rather different employment transitions, only because of similarity in earnings or demographics. More sophisticated versions of the two-step GFE circumvent this contamination issue, but we believe that putting them into practice would be beyond the scope of this paper.

[2] We could have followed different methods to estimate our model. As in Ahn and Hamilton (2020) and Hall and Kudlyak (2019), we could have estimated type-specific parameters and a distribution of latent types via maximum likelihood. Alternatively, we could have estimated individual-specific parameters via maximum likelihood based on the employment history of each worker. The dimension of our dataset—both in terms of number of individuals and number of observations per individual—makes these methods computationally cumbersome. In contrast, the two-step GFE method is computationally light even on our large and long dataset. Moreover, given the dimension of our dataset, we can appeal to the consistency properties of the two-step GFE estimates.

federal government, self-employment, contracting work, or any other form of employment that is not covered by Unemployment Insurance.[3]

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.

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 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 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.[4] 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.

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

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

We then 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 (which includes imputed spells), between 1 and 4 quarters, and between 4 and 8 quarters; (iii) 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 (i) tells us the distribution of job durations for an individual. Statistic (ii) tells us the distribution of unemployment durations for an individual. Statistic (iii) tells us how much time the individual spends in unemployment. Statistic (iv) together with (iii) tells us about direct job-to-job transitions.

Obviously, there are other ways to summarize the pattern of employment transitions of an individual. The particular statistics that we chose are called for by the properties of the theoretical model that we are going to estimate. In our model, each job spell of a particular worker has a duration that is drawn independently from the same distribution. Similarly, each unemployment spell of a particular worker has a duration that is drawn independently from the same distribution. Because of this "memory-less" property of the model, we summarize the history of an individual worker through the duration distribution of job spells and unemployment spells. For similar reasons, we summarize direct job-to-job transitions with the average number of jobs and with the fraction of time spent in unemployment.

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. 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.

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, $n = 1, 2, ...$, 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[5] 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 remarkably different patterns of labor market transitions. Consider the distribution of unemployment spell durations for different types. For αs, the fraction of unemployment spells lasting less than 1 quarter is 79% and the fraction of spells

[5]Other criteria for choosing the number of types (e.g., the elbow criterion) also return $J = 3$.

	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
Annualized Earnings (2014$)	$52,464	$35,021	$25,496

Table 1: Descriptive statistics for each worker type

lasting more than 1 year is only 5%. For βs, 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 γs, 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, αs typically have short unemployment spells and very rarely have long ones; βs are less likely to have short unemployment spells but they also have very few long ones; γs are much more likely to experience long unemployment spells compared to αs and βs. The average time spent in unemployment is 3.5% for an α, 9.6% for a β, and 29.2% for a γ.

Next, consider the distribution of job durations for different types. For αs, 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 βs, 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 γs, 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, αs are 50% more likely to remain in the same job for more than 2 years than βs, and 4 times more likely than γs. Conversely, γs are 70% more likely to leave a job within 1 quarter than βs, and 3 times more likely than αs.

Overall, our classification of workers into types paints a clear picture. When unemployed, αs are likely to find a job quickly and, once they find it, they are likely to keep it for more than 2 years. Unemployed βs find a job less quickly than αs and, once they find it, they are less likely to keep it for more than 2 years. Unemployed γs 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 γs 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: αs make about 1.5 times what βs earn and βs earn about 1.4 times as much as a γ. Follow-up studies that apply our clustering strategy to administrative data from Italy (Spinella 2022), Denmark (Darougheh and Lundgren 2022) and Canada (Castro, Lange and Poschke 2024) paint a very similar picture.

Notice that our classification builds on the implicit assumption that a worker's type is fixed. In order to validate this assumption, we reclassify workers based on their pattern of employment transitions over the pre-Great Recession period of 1997-2008. We find that the fraction of workers that are classified in the same way when we use the full period 1997-2014

	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

as when we use the subperiod 1997-2008 is very high (approximately 90%). Similarly, Spinella (2022) finds a very high autocorrelation in the classification of workers across non-overlapping periods.

2.4 Types and observables

It turns out that a worker's type—a summary of the worker's pattern of employment transitions—cannot be explained by a 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 α 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 more likely to be an α, a non-white individual is less likely to be an α, 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 that are used for clustering. 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 αs, 0.6% for βs, and 5% for γs.

Table 3 reports the fraction of αs, βs and γs in different demographic and industry groups. The first two rows show that the fraction of αs, βs and γs are 62%, 25% and 13% among

	α-workers	β-workers	γ- workers
Race			
White	61.7%	25.4%	13.0%
Non-white	54.2%	28.8%	17.0%
Gender			
Male	58.9%	26.2%	14.9%
Female	58.9%	27.1%	14.1%
Education			
< High school	49.0%	30.4%	20.6%
High school	56.8%	27.3%	15.9%
Some college	59.7%	26.4%	13.9%
College+	65.4%	24.1%	10.4%
Industry			
Retail	55.6%	27.8%	16.7%
Services	49.1%	28.8%	22.1%
Manufacturing	64.5%	25.5%	10.0%
Skilled	65.3%	24.5%	10.2%
Education	63.1%	27.4%	9.5%
Health	62.3%	25.9%	11.9%
Admin.	43.4%	27.3%	29.3%
Age			
20 - 39	55.1%	26.5%	18.4%
40 - 69	61.5%	26.9%	11.6%

Table 3: Type composition of demographic groups

white workers, and 54%, 29% and 17% among non-white workers. There are some differences in the composition of types among the two groups, but the differences are relatively small. In both groups, the majority of workers are αs. In both groups, there are between 10 and 20% of γ-workers. Similar observations apply to the composition of worker types among workers with different levels of education. The fraction of αs among workers with a college degree is 65%, but it is close to 50% even among workers that have not completed high school. The fraction of γs is 20% among high school dropouts, but it is still 10% among college graduates. With respect to industry, the highest fraction of αs is 64% (manufacturing) and the lowest is 43% (administration). The highest fraction of γs is 29% (administration) and the lowest is 9% (education). Again, there are differences in the composition of types across industries, but αs are the dominant type in every industry and γs are present in every industry. There are also some small differences in the composition of worker types by age.

The fact that demographic characteristics and industry cannot predict a worker's type may come as a surprise. It is well documented that the unemployment rate and the transition rates are different for high school and college graduates, whites and non-whites, younger and older workers. Indeed, Table 3 shows that the composition of types varies by education, race and age. Yet, Table 2 shows that education, race and age cannot predict whether a worker is an α, β or γ. One way to understand these seemingly contradictory observations is through an analogy with Mincerian wage regressions. Average wages are different for high school graduates and college graduates, whites and non-whites, younger and older workers. Yet, Mincerian wage regressions show that observable characteristics account for a very small fraction of the overall wage dispersion. This happens because the difference in the average wages of workers with different observables are small compared to the difference in wages among workers with the same observables. Tables 2 and 3 reveal that the same is true for the pattern of workers' transitions.

The fact that demographics do not predict a worker's type is important, as it implies that, in order to capture the full extent of heterogeneity in the pattern of workers' transitions, one needs to group workers based on latent types rather than on observable characteristics. Indeed, if we estimated fixed-effects that are, say, education-specific rather than type-specific, we would underestimate the extent of heterogeneity in workers' transitions. In turn, we might understate the extent to which changes in the composition of the pool of unemployed workers contribute to the cyclical fluctuations of the UE rate, as well as understate the extent to which changes in the composition of the pool of employed workers contribute to the cyclical fluctuations of the EU and EE rates.

The fact that demographics do not predict a workers' type leaves us with a puzzle: Who are α, β and γ workers?[6] At a superficial level, our estimated structural model implies that αs, βs and γs are workers who, essentially, differ with respect to their ability to locate jobs, their baseline productivity, and their distribution of match-specific productivities across different jobs. At a deeper level, these differences may originate from differences in personality, taste or circumstances that are largely orthogonal to demographic characteristics. For instance, a worker's type may reflect the individual likelihood to be willing to follow the instruction of management and cooperate with coworkers. According to this interpretation, αs are "gregarious" types, who interact well with different groups of people and, hence, are more productive on average and equally productive in all jobs. As a result, αs find jobs more quickly and stick with them for longer. In contrast, γs are "difficult" types, who only interact well with specific groups of people and, hence, are less productive on average, unproductive in some jobs and very

[6]We are grateful to Mike Pries for inspiring much of the content of this paragraph.

productive in others. As a result, γs find jobs more slowly and are unlikely to stick with them. Another possible interpretation is that a worker's type reflects the individual taste for change or stability. In this view, αs desire stability, and for this reason they are eager to perform well in all jobs. In contrast, γs enjoy change, and for this reason they rather quit a job that accommodate the demands of management and coworkers.

Obviously, we are not the first to measure the extent of latent heterogeneity in the pattern of workers' transitions across employment states. Using the CPS panel data, Hall and Kudlyak (2019) use individual transitions to estimate, via maximum likelihood, a distribution of workers' types and, for some of the types, a vector of transition rates. They identify 5 types of workers: always employed, always out of the labor force, often employed, often unemployed, and often out of the labor force. Since the CPS only tracks individuals for a year and a half, about 70% of individuals do not experience any transitions over the period of observation. As a result, they find that more than 50% of individuals belong to the "always employed" type and, for this type, they have no estimates of transition rates. Thus, the findings in Hall and Kudlyak (2019) are not comparable to ours. Shibata (2015) uses the CPS panel data to estimate an initial distribution of workers across unobservable "attachment" and employment states, and transition rates across these states. Shibata (2015) finds, like we do, that unobservable heterogeneity is needed to fit the data. He does not provide, however, a theory of transitions across attachment and employment states and, hence, cannot produce counterfactuals. Ahn, Hobijn and Sahin (2023) use the CPS to estimate a Hidden Markov Model to identify different types of workers and track their composition over time. As in Shibata (2015), the analysis is empirical and cannot produce counterfactuals.

Ahn and Hamilton (2020) use the time-series of the cross-sectional distribution of unemployment durations to estimate, via maximum likelihood, the inflow of two types of workers into the pool of unemployment and, for each type, a time-varying UE rate. Consistently with our findings, Ahn and Hamilton (2020) find large differences in the UE rates of the two types. Ahn and Hamilton (2020), however, do not provide estimates for the EU and EE rates of the two types, which, as documented in Table 1, is an important dimension of heterogeneity. The same comment applies to the literature that estimates workers' heterogeneity in the UE rate from repeated unemployment spells (e.g., Alvarez, Borovickova and Shimer 2018).

Morchio (2020) uses panel data from the NLSY to classify workers based on the amount of time spent in unemployment during prime age. The first type of worker includes individuals at the top 10% of the distribution of time spent in unemployment, the second type of worker includes everyone else. He finds that the first type has both a lower UE rate and a higher EU rate than the second type. In contrast to Morchio (2020), we use the entire pattern of employment transitions to classify workers, and we find that the time spent in unemployment is not a sufficient statistic for identifying a worker's type. Karahan, Ozkan and Song (2019) use annual data from the Social Security Administration to classify workers based on their decile of lifetime earnings and, conditional on lifetime earnings, they measure differences in employment transitions. In stark contrast to Karahan, Ozkan and Song (2019), we classify workers based on employment transitions.

3 Modeling Heterogeneity

In this section, we develop an equilibrium model of workers' transitions across employment states based on Menzio and Shi (2011). Firms spend resources to open vacancies which advertise the terms of trade offered to hired workers. 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, his baseline productivity, and the distribution of his productivity across different matches.

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 $i = 1, 2, \dots I$. 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. [7] 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 $q(\infty) = 0$.

[7] 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).

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 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 bilaterally efficient contract that is worth v in lifetime income.[8] 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.

[8]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 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.

At the *production* stage, a worker of type i who is unemployed receives an income equal to b_i units of output. 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.

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.[9] 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

In general, a Recursive Equilibrium of the labor market would be such that the value and policy functions depend on the aggregate state of the economy $\psi \equiv \{x, u_i, n_i, g_i\}$, which includes a large vector describing the distribution of workers across types and employment states. For this reason, solving for a Recursive Equilibrium would be, in general, computationally challenging anywhere outside of the steady state. As established in Menzio and Shi (2011), however, the assumption of directed search guarantees that the unique Recursive Equilibrium is block recursive, i.e. the equilibrium is such that the value and policy functions depend on the aggregate state of the economy ψ only through the realization of the aggregate component of productivity x and not through the entire distribution of workers across types and employment states $\{u_i, n_i, g_i\}$. For this reason, solving for the equilibrium is as computationally easy in and out of steady state.

To formally define an equilibrium, let us introduce a few additional pieces of notation. Let $U_i(x)$ denote the lifetime income for a worker of type i who is unemployed at the beginning of the production stage. Let $\tilde{V}_i(x)$ 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 $V_i(z, x)$ 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 $\theta_i(v, x)$ denote the equilibrium tightness of submarket $m = \{v, i\}$.

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

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

In the current period, the worker's income is b_i. In the next period, the worker finds a job with probability $\lambda_u^i p(\theta_i(v, \hat{x}))$. In this case, the worker's continuation value is v. The worker does

[9] 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 have let the probability λ_u^i depend on the worker's type, while keeping $k_i = k$ for all types of workers. The two approaches are essentially equivalent.

not find a job with probability $1 - \lambda_u^i p(\theta_i(v, \hat{x}))$. In this case, the worker's continuation value is $U_i(\hat{x})$.

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

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

In the current period, the sum of the worker's income and firm's profit is $x y_i z$, 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 $U_i(\hat{x})$ and the firm's continuation value is 0. The worker moves from the current job to a new job with probability $(1 - d)\lambda_e^i p(\theta_i(v, \hat{x}))$. 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 $(1 - d)(1 - \lambda_e^i p(\theta_i(v, \hat{x})))$. In this case, the firm's and worker's joint continuation value is $V_i(z, \hat{x})$. 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 $\tilde{V}_i(x)$ of a match of unknown quality between a firm and a worker of type i is given by

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

In the current period, the expected output of the match is $x y_i$. 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 $U_i(\hat{x})$. The worker searches on-the-job and finds a new job with probability $(1 - d)\lambda_e^i p(\theta_i(v, \hat{x}))$. In this case, the joint continuation value is v. The worker and the firm remain together with probability $(1 - d)(1 - \lambda_e^i p(\theta_i(v, \hat{x})))$. In this case, the joint continuation value is $V_i(z, \hat{x})$. The firm and the worker do not learn the quality of their match with probability $1 - \phi_i$.

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

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

and $\theta_i(v, x) \geq 0$, 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, $q(\theta_i(v, x))$, times the firm's value from filling a vacancy, $\tilde{V}_i(x) - v$, 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

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

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

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

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

Notice that, for all $\theta \geq 0$, there exists a v such that $\theta_i(v, x) = \theta$. 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 $\theta = \theta_i(v, x)$. 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

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

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, $\tilde{V}_i(x) - v_0$, 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

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

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, k_i, of searching in a submarket with a marginally higher tightness must be equal to the benefit, $p'(\theta)(\tilde{V}_i(x) - v_0)$. 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 $\theta^*_{i,u}(x)$ the optimal search strategy for a worker who is unemployed. That is, $\theta^*_{i,u}(x)$ is the solution to (3.8) for $v_0 = U_i(x)$. We denote as $\theta^*_{i,e}(z, x)$ the optimal search strategy for a worker who is employed in a match of known quality z. That is, $\theta^*_{i,e}(z, x)$ is the solution to (3.8) for $v_0 = V_i(z, x)$. Since $V_i(z, x)$ is strictly increasing in z, $\theta^*_{i,e}(z, x)$ is strictly decreasing for all $z < Q_i(x)$ and zero for all $z \geq Q_i(x)$, where $Q_i(x)$ is defined as the match quality such that $V_i(Q_i(x), x) = \tilde{V}_i(x) - k_i/p'(0)$. 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

$$U_i(x) \lesseqgtr v_0 + \lambda^i_e D_i(v_0, x). \tag{3.9}$$

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 $d_i^*(z, x)$ the optimal separation probability for a firm and a worker in a match of known quality z. That is, $d_i^*(z, x)$ denotes the optimal separation probability for $v_0 = V_i(z, x)$. Since $V_i(z, x)$ is strictly increasing in z, there exists a reservation quality $R_i(x)$ such that $d_i^*(z, x) = 1$ for all $z < R_i(x)$ and $d_i^*(z, x) = \delta_i$ for all $z \geq R_i(x)$. Similarly, we denote as $\tilde{d}^*(x)$ 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.[10] An unemployed worker finds a job with probability $\lambda_u^i p(\theta_i(v, x))$. As long as the worker does not observe the quality of his match, he moves into unemployment with probability $\tilde{d}^*(x)$ 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 $z < R_i(x)$. If $z \in [R_i(x), Q_i(x))$, 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, and he searches in submarkets with a lower tightness the lower is z. If $z \geq Q_i(x)$, the worker moves into unemployment only for exogenous reasons and does not search for a better job—in the sense that he searches in a submarket with zero tightness.

Notice that equilibrium workers' transitions across employment states have a special feature. When the aggregate component of productivity x is constant, the duration of each unemployment spell for each worker of type i is an independent draw from a common probability distribution. Similarly, when x is constant, the duration and destination (unemployment or employment at another job) of each job spell for each worker of type i are an independent draw from a common probability distribution. This "memoryless" property of the equilibrium employment transitions is what motivated our choice of the statistics used to summarize workers' heterogeneity.

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

[10] In a Block Recursive Equilibrium, the value and policy functions—and the implied workers' employment transition probabilities—are independent from the laws of motion for the distribution of workers across types and employment states. For this reason, we relegated the formulation of these laws of motion to Appendix B.

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.[11]

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.[12]

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, which we normalize to 1; (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.[13]

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.[14] 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 αs, 15% for βs, and

[11] As we match moments observed in the data over the pre-Great Recession period 1997-2008 rather than over the entire period between 1997 and 2014, there are some differences between the moments reported in this section and those reported in Section 2.

[12] Menzio and Shi (2011), 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.

[13] 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.

[14] 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.

10% for γs.[15] 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 rate of workers of type i in the period preceding the Great Recession, i.e. an unemployment rate of 4.2% for αs, 12.5% for βs, and 28.8% for γs. 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_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 αs, this fraction is 21% (about one-half of matches that last no more than 24 months). For βs, it is 18% (about one-third of matches that last no more than 24 months). For γs, 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, through dynamic selection, the shape of the tenure distribution.

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

[15] 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 λ_u^i to fall with the duration of an unemployment spell. We decided against this generalization in order to keep the model simpler.

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 4: Model parameters

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 4 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 (solid line), the implied steady-state distribution of workers across match-specific quality (dashed line), 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 αs, the calibrated distribution is a Weibull with shape 4.5 and scale 0.25. Such distribution 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 βs, 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 γs, the calibrated distribution is

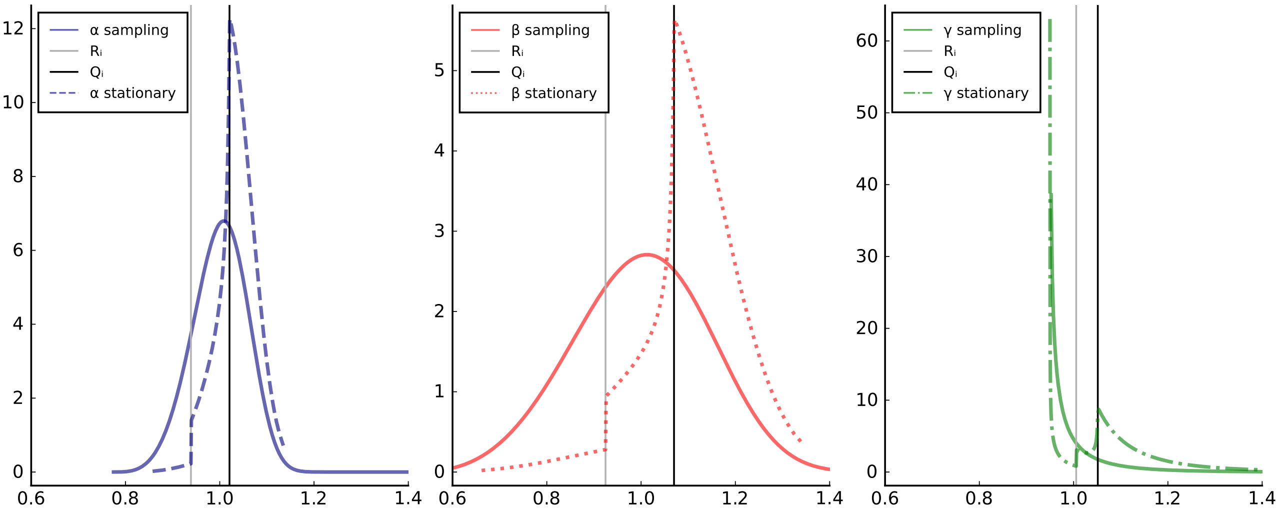

Figure 2: Match quality distributions $f_i(z)$ (solid lines) with the stationary distribution of match quality (dotted/dashed lines) by worker type.

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 5. 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, which implies that the probability that a worker leaves a job as soon as its quality is revealed is low. The small right tail implies that the return to searching for better matches is low and, hence, Q_i is close to R_i. This means that the probability that a worker keeps a job until he is forced into unemployment for exogenous reasons is high. 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—unless they are employed in a match with a quality at the top 15% of the distribution. The right-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.

Types also differ with respect to baseline productivity and unemployment income. For αs, 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 βs, 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 γs, 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

	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 5: Employment duration moments

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.

Lastly, let us comment on how types differ with respect to the parameters describing the search process. The cost of maintaining a vacancy for αs is equal to 2.8 units of output. Also, for αs, the probability of 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 βs, 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 γs, 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.

The calibration outcomes provide an interpretation of the empirical differences in the pattern of employment transitions for different types of workers. Workers of type α have an average productivity that is high relative to their unemployment income and, hence, they enjoy 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 an average productivity that is low relative to their unemployment income and, hence, they enjoy 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 γs having a low probability of remaining on a job for a long period of time. Overall, the search process of αs—which entails finding any job—is fast. The search process of γs—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. In Section 5.2, we examine the predictions of the theory with respect to the relationship between the average UE rate and unemployment duration, and the relationship between the composition of the unemployment pool and unemployment duration.

5.1 Earning losses of displaced workers

It is well-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 that 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 an important 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 the capital after losing it. The test is even more important for our theory, which posits that different types of workers follow very different search processes.

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 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 αs (dashed line), βs (dotted line) and γs (dash-dotted line). The average earnings losses for displaced workers are sizeable and quite persistent.[16] 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 αs, earnings losses are smaller and more transitory than average (20% after six quarters, and 10% after fourteen). For γs, earnings losses are much larger and much more persistent than average (50% after six quarters, and still about 50% after fourteen quarters). For βs, earnings losses are close to the

[16] 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.

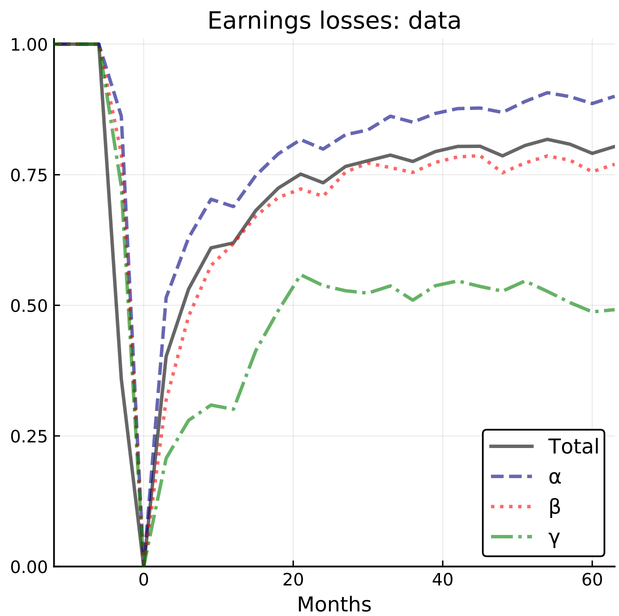

Figure 3: Earnings losses from job separation in steady state.

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.[17]

The theory provides a simple explanation for the magnitude and persistence of the earning losses for different types. Consider an α 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 αs 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 αs have about a 50% chance of sampling a job with quality $z \geq Q$. Overall, once an α is displaced from a high-tenure job, he is likely to quickly find a new job with a quality that is similar to the quality of his old job. Now consider a γ 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 γs 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 γs 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 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.

The fact that it succeeds in reproducing the earnings losses of displaced workers is an actual test of the theory, and not a simple and mechanical implication of how the theory is calibrated. The earnings losses of displaced workers depend both on the duration distribution of job and unemployment spells—objects that are targeted in the calibration—but also on the workers' wages before and after displacement—objects that are not targeted in the calibration.

[17] In Appendix C, we report the analogue of Figure 3 for the Great Recession. We find that the earning losses are larger and especially so for γ's. We find that the theory correctly predicts larger earning losses for all types of workers, and especially for γ's.

Indeed, our calibration relies entirely on the theory in order to map the observed job duration distribution into a job quality distribution and, in turn, wages. Therefore, the fact that the theory succeeds in reproducing the earning losses validates the theoretical linkages between the quality of a job, the duration of a job, and the worker's wage. The fact that the theory succeeds in reproducing the earnings losses of displaced workers is also somewhat surprising, since matching the magnitude and persistence of earnings losses is known to be a challenge for basic search-theoretic models of the labor market and typically requires resorting to human capital depreciation, stigmatization, or other sources of scarring[18] (see, e.g., Davis and von Wachter 2011, Jarosh and Pillosoph 2019).

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). Reproducing the relationship between the UE rate and duration is a useful test for 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 on the UE rate of different types (i.e. true duration dependence). Since our theory is calibrated to the distribution of unemployment and job spell durations but not to the number of unemployment spells, it does not mechanically reproduce the composition of types at the beginning of an unemployment spell. Since our theory rules out true duration dependence, it does not mechanically reproduce the decline in the UE rate and the change in composition of the unemployment pool.

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 4, 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 αs, 35% of βs, and 25% of γs. After one year of unemployment, the pool has 25% of αs, 35% of βs, and 40% of γs.

In the bottom panel of Figure 4, 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

[18] 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.

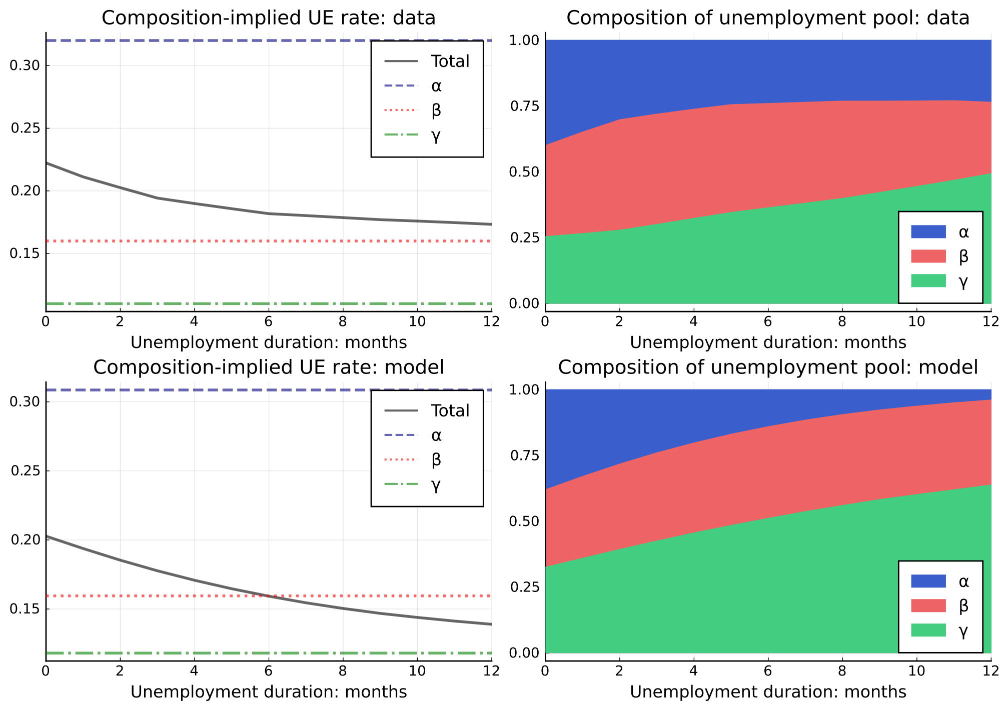

Figure 4: Left panels: type-specific and aggregate UE rates by duration; right panels: composition of unemployment pool by duration. Top panels are from the data and bottom panels are from the model.

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 αs, 30% of βs, and 30% of γs. After one year of unemployment, the theory predicts that the pool of unemployment contains 10% of αs, 40% of βs, and 50% of γs. 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.

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 αs, 15% for βs and 10% for γs. 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 entirely 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

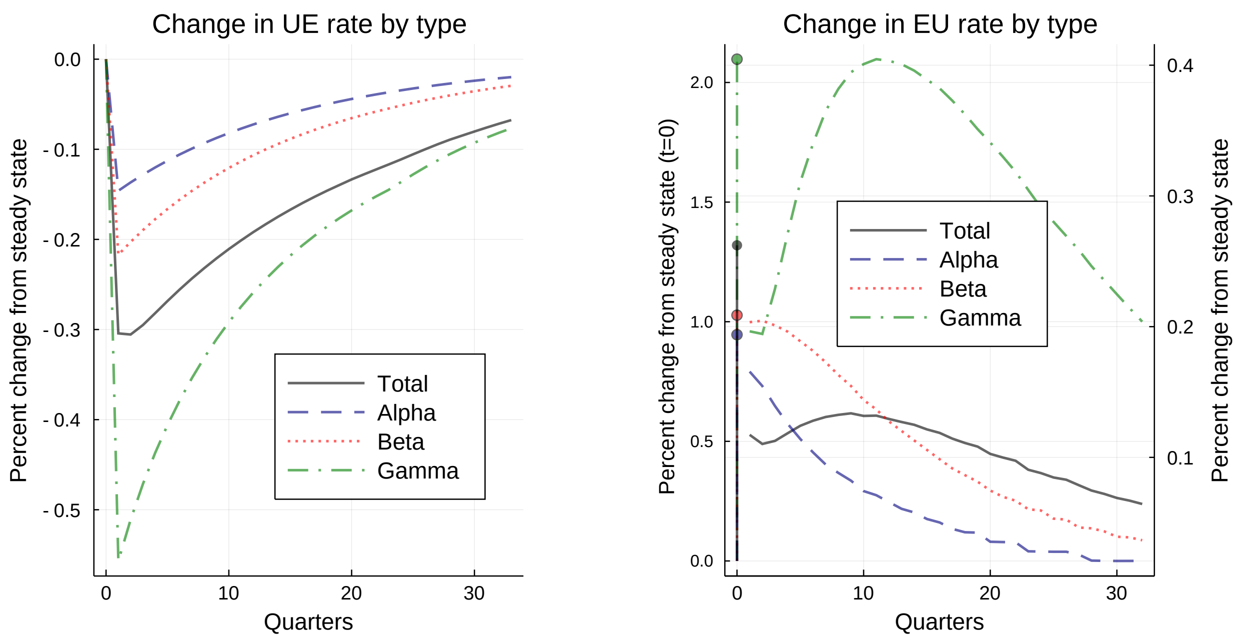

Figure 5: Left panel: model-implied UE rates by type; right panel: model-implied EU rates by type (left axis refers to the response on impact, right axis refers to everything thereafter)

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 the extent of heterogeneity in the UE rate by estimating a random-effect model on a panel of unemployment spell durations. They find that heterogeneity accounts for only a fraction of the decline in the average UE rate. In contrast, we recover the extent of heterogeneity in the UE rate using the entire pattern of employment transitions of an individual and the behavior of similar workers.

6 Macro Measurement

In this section, we turn our attention to the theory's predictions with respect to labor market fluctuations. In Section 6.1, we use the theory to measure the impact of a negative shock to the aggregate component of productivity on type-specific and aggregate labor market outcomes. In Section 6.2, we contrast the predictions of the theory with those that would emerge from a version of our model with either 1 or 2 types of workers. In Section 6.3, we compare the predictions of the theory with the empirical behavior of αs, βs and γs during and after the Great Recession of 2008-2009.

6.1 Aggregate productivity shocks

We first want to use the theory to measure the effect of an aggregate productivity shock on the labor market. Specifically, we take the labor market at its non-stochastic steady-state and measure its response to a one-time negative shock to the aggregate component of productivity x with a magnitude of 10% and a half-life of 3 years.

The left panel of Figure 5 plots the response of the UE rate to the aggregate productivity shock for different types of workers. On impact, the UE rate declines for all types of workers. Intuitively, the decline in the aggregate component of productivity lowers the expected gains from trading in the labor market and, for this reason, it lowers the tightness of the submarket in which unemployed workers look for a job. The magnitude of the initial decline of the UE rate is different for different types of workers. The initial decline of the UE rate is 15% for αs, by 23% for βs, and by 60% for γs. Intuitively, the difference in the decline of the UE rate for different type of workers is due differences in the size of the gains from trading in the labor market. For αs, the gains from trade are large, since their output while employed, xyz, is large relative to their income while unemployed, b. For γs, the gains from trade are small, since their income while employed is small relative to their income while unemployed. As a result, the aggregate productivity shock reduces the gains from trade for αs by a smaller proportion than γs and, in turn, it leads to a smaller decline in the UE rate for αs than for γs. The speed at which the UE rate recovers is, however, the same for all types of workers, and it is equal to the speed at which the aggregate productivity shock recovers.

The right panel of Figure 5 plots the response of the EU rate to the aggregate productivity shock for different types of workers. On impact, the EU rate increases for all types of workers. Intuitively, the shock to x reduces the value of employment relative to the value of unemployment for all types of workers 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 αs. 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 that are discovered to be below the new and higher reservation quality. The effect of the shock dissipates very quickly. Now, consider γs. 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 slowly.

Let us explain why the response of the EU rate is different for different types of workers. On impact, the effect of the productivity shock on the EU rate depends on the density of the cross-sectional distribution of employed workers around the reservation quality. In turn, the density of the cross-sectional distribution of employed workers around the reservation quality depends on the stock of relatively new matches (whose quality is drawn from f) relative to the stock of relatively old matches (whose quality first order-stochastically dominates the truncation at R of f). Since γs are less likely to keep a job than αs, the ratio of new to old matches is higher for γs and, in turn, the density of the cross-sectional distribution of employed workers around the reservation quality tends to be higher for γs. After the impact, the effect of the shock depends on the search process of displaced workers. The persistence is higher for γs than for αs because displaced γs typically go through about 7 EU transitions before finding a stable match, while displaced αs go only through 2 EU transitions.

The left panel in Figure 6 plots the response of the unemployment rate for different types of workers. For αs, the unemployment rate increases by 2 percentage points and is re-absorbed quickly (half-life of about 1 year). For βs, the unemployment rate increases by 5 percentage points and is re-absorbed more slowly (half-life of about 3 years). For γs, the unemployment rate increases by 20 percentage points and is re-absorbed very slowly (half-life 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

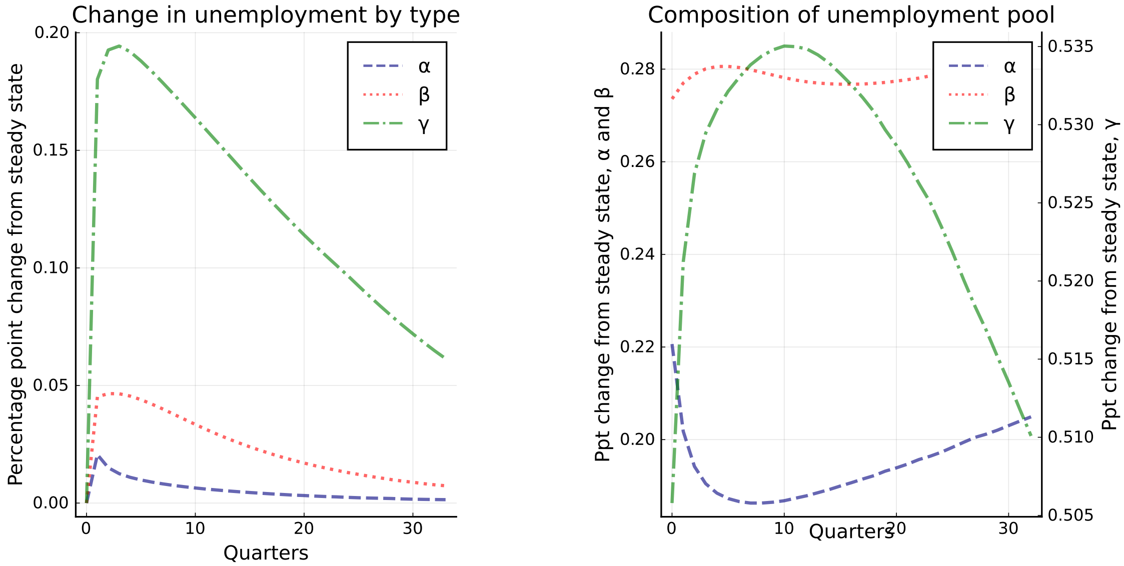

Figure 6: Model-implied unemployment rates and composition in response to a 10% decline in aggregate productivity

the response of the type-specific UE and EU rates.

The right panel in Figure 6 plots the change in the composition of the unemployment pool. At the non-stochastic steady state, the unemployment pool has about 22% αs, 28% βs, and 50% γs. On impact, the shock increases the unemployment of γs slightly more than proportionally than the unemployment of αs. As a result, on impact, the composition of the unemployment pool tilts slightly towards γs and away from αs. Over time, the excess unemployment of αs is re-absorbed quickly, while the excess unemployment of γs is re-absorbed slowly. As a result, the composition of the unemployment pool tilts further towards γs and away from αs and peaks after about 10 quarters. After that, the composition of the unemployment pool starts to revert back towards the steady-state.

The changing composition of the unemployment pool helps us understand the aggregate response of the labor market to the aggregate productivity shock. The black solid line in the left panel of Figure 5 plots the response of the aggregate UE rate. On impact, the aggregate UE rate declines by about 30%. This is slightly higher than the average of the decline in the UE rate for different types weighted by the steady-state composition of the unemployment pool because of the initial effect of the shock on the composition of the unemployment pool. Over time, the decline in the aggregate UE rate dissipates, but more slowly than any of the type-specific UE rates. 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 also caused by the effect of the shock on the composition of the unemployment pool. As the composition of the unemployment pool tilts further towards γs, the weight on the UE rate of these workers increases and it slows down the recovery of the aggregate UE rate.

The black solid line in the right panel of Figure 5 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 αs and βs that were displaced by the shock return to stable matches, while the γs that were displaced by the shock experiment and then leave several matches. Hence, over time, the aggregate

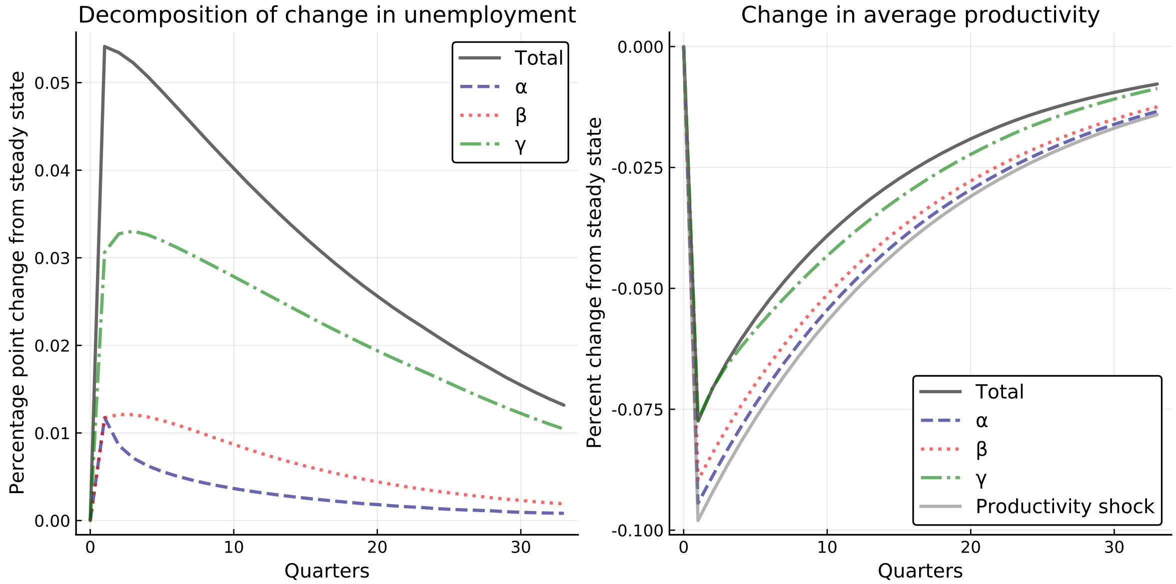

Figure 7: Left panel: decomposition of the response of the aggregate unemployment rate; right panel: aggregate and type-specific productivity of labor

response of the EU rate becomes entirely driven by γs.

The aggregate behavior of unemployment and labor productivity is displayed in Figure 7. On impact, the aggregate unemployment rate increases by 5.4 percentage points, with 25% of the increase due to the rise in the unemployment of αs, 21% due to the rise in the unemployment of βs, and 54% due to the rise in the unemployment of γs. The increase in the unemployment of αs and βs is re-absorbed much faster than the increase in the unemployment of γs. Hence, the excess aggregate unemployment rate is eventually entirely due to the excess unemployment of γs and, like the excess unemployment of γs, it dissipates slowly. The half-life of the increase in the aggregate unemployment rate is 5 years.

Lastly, we turn to labor productivity (i.e. output per employed worker). On impact, labor productivity falls, but by less than the aggregate component of productivity does. Specifically, 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. Specifically, 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 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 γs, who have the lowest productivity among all types. 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 time, the displaced workers who first find a new stable match are αs and βs. Hence, over time, the composition of the employment pool tilts further towards αs and βs and further away from γs. Since αs and βs are more productive than γs, labor productivity recovers faster than the underlying shock.

Let us summarize our findings. First, the aggregate unemployment rate is sensitive to the negative shock to aggregate productivity. Specifically, the elasticity of the aggregate unemployment rate with respect to the aggregate productivity shock is about 7, and the elasticity of the aggregate unemployment rate with respect to average labor productivity is about 10. In this sense, the model implies a large amplification of productivity shocks. Second, the aggregate unemployment rate recovers more slowly than productivity does. Specifically, the half-life of the increase in the aggregate unemployment rate is 5 years, while the half-life of the productivity shock is 3 years, and the half-life of average labor productivity is 2 years. In this sense, the model implies significant propagation of productivity shocks and, for this reason, it creates a "jobless recovery". Third, the response of labor productivity is muted relative to the underlying productivity shock. Specifically, labor productivity declines less and recovers more quickly than the productivity shock. As we will detail in Section 6.3, the response of the labor market to a negative aggregate shock is quantitatively similar to the actual behavior of the labor market during and after the Great Recession.

The response of the labor market to a negative aggregate shock is very different in our model than in the textbook search-theoretic model of the labor market (e.g., Pissarides 1985). First, in the textbook model, aggregate productivity shocks generate unemployment fluctuations that are an order of magnitude smaller than in our model (e.g., Shimer 2005). The literature has identified several ways to increase the responsiveness of unemployment to productivity shocks: sticky wages (e.g., Hall 2005, Gertler and Trigari 2009, Kennan 2010, Menzio and Moen 2010, Fukui 2021, Menzio 2022); small gains from trading in the labor market (e.g., Hagedorn and Manovskii 2008, Sargent and Lijndquivst 2017); heterogeneous match quality (e.g., Menzio and Shi 2011). In our model, unemployment is very responsive to aggregate shocks because of γs—a small fraction of workers whose UE rate is very sensitive to productivity shocks because they have relatively small gains from trade (yet, not as small as in Hagedorn and Manovskii 2008[19]), and whose EU rate is very sensitive to productivity shocks because many of them are in matches close to the reservation quality (but more so than in Menzio and Shi 2011). Quantitatively, these two amplification channels are equally important.[20]

Second, in the textbook model, unemployment fluctuations have the same persistence as the underlying aggregate productivity shocks. Mechanisms that are known to increase the persistence of unemployment fluctuations 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 model, unemployment fluctuations are much more persistent than the underlying productivity shocks because of γs—a small fraction of workers whose unemployment recovers very slowly after a negative productivity shock because their search process is slow, as it does not simply involve finding a job, but finding one of the few jobs in the right tail of the quality distribution. Hence, in our model, the persistence of unemployment is generated by the same mechanism highlighted in Pries (2004).[21]

[19]In our model, the ratio of unemployment income to average labor productivity is 90% for γs. In Hagedorn and Manovskii (2008), the ratio is 95.5%. The difference is significant because the elasticity of the UE rate with respect to productivity shocks diverges to infinity as the ratio of unemployment income to labor productivity approaches 1.

[20]A version of the model without match quality heterogeneity and, hence, without an endogenous job-destruction margin, generates an elasticity of the unemployment rate relative to labor productivity that is approximately half as large as in the baseline model.

[21]Not every model in matches are experience goods generates much propagation of negative productivity shocks. In Menzio and Shi (2011), the match-quality distribution is calibrated to reproduce the cross-sectional distribution of job durations. Since not many jobs are destroyed within their first year, the match-quality distribution is such that a relatively small fraction of matches is below the reservation quality. Therefore, it does not take long for a displaced worker to find stable employment and, hence, a negative productivity shock does not generate much propagation. The same holds true for α-workers in our model. In contrast, γ-workers

Third, in the textbook model, the correlation between unemployment and labor producitivty is nearly perfect, which runs against recent empirical evidence (e.g., Gali and Van Rens 2017). In order to address the low correlation between unemployment and productivity, the literature has 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 model, the correlation between unemployment and productivity is weakened by the fact that unemployment is more persistent than productivity shocks (due to the behavior of gammas), and by the fact that labor productivity is less persistent than productivity shocks (due to composition effects).

6.2 The role of heterogeneity

When discretizing the heterogeneity of workers with respect to their pattern of transitions across employment states, we found that heterogeneity was best summarized by having three different types of workers. For this reason, our baseline model, which we shall refer to as the 3T model, had three types of workers. It is, however, unclear whether workers' heterogeneity is necessary to study macroeconomic phenomena.

In order to assess the role of heterogeneity in shaping the macroeconomic response of the labor market to an aggregate productivity shock, we consider two alternative versions of the baseline model. The first alternative model, which we refer to as the 1T model, features a single type of worker. That is, in the 1T model, we collapse all types into a representative type. The 1T model is calibrated by matching the appropriately weighted moments of the type-specific moments used to calibrate the baseline model. As we shall see, aggregate productivity shocks have significantly smaller and more transient effects on the aggregate labor market in the 1T than in the 3T model, thus revealing the importance of accounting for some worker's heterogeneity to analyze macroeconomic phenomena. The second alternative model, which we refer to as 2T, features two types of workers. One type is calibrated to the average moments of α and β workers. The other type is calibrated to the moments of γ workers. As we shall see, aggregate productivity shocks have nearly identical effects on the labor market in the 2T model as in the baseline model. This reveals that, for macroeconomic analysis, it is sufficient to account for the difference between γs and other workers.

In Figures 8 and 9, we report the response of the 1T model to a 10% negative shock to the aggregate component of productivity with a half-life of 3 years. In response to the shock, the UE rate falls by 18% with a half-life of 3 years. In contrast, the average UE rate in the 3T model falls by 30% with a half-life of 4 years. The difference in the magnitude of the response in the UE rate is due to the fact that the response of the UE rate is convex in the workers' gains from trade. Hence, the UE rate declines more in a model where heterogeneous workers have different gains from trade than in a model where a representative worker has average gains from trade. The difference in the half-life of the response of the UE rate is due to the fact that, in a model with a representative worker, the composition of the unemployment pool does not change over time and, hence, the UE rate simply tracks the productivity shock.

In response to the productivity shock, the EU rate in the 1T model increases by about 50% on impact. Afterwards, the EU rate increases by 10% with a half-life of 3 years. In contrast, the average EU rate in the 3T model increases by about 200% on impact and, afterwards, increases

have a 70% probability of leaving their job within 1 year and, for this reason, the fraction of their matches below the reservation quality is very high and, in turn, the propagation mechanism is very strong.

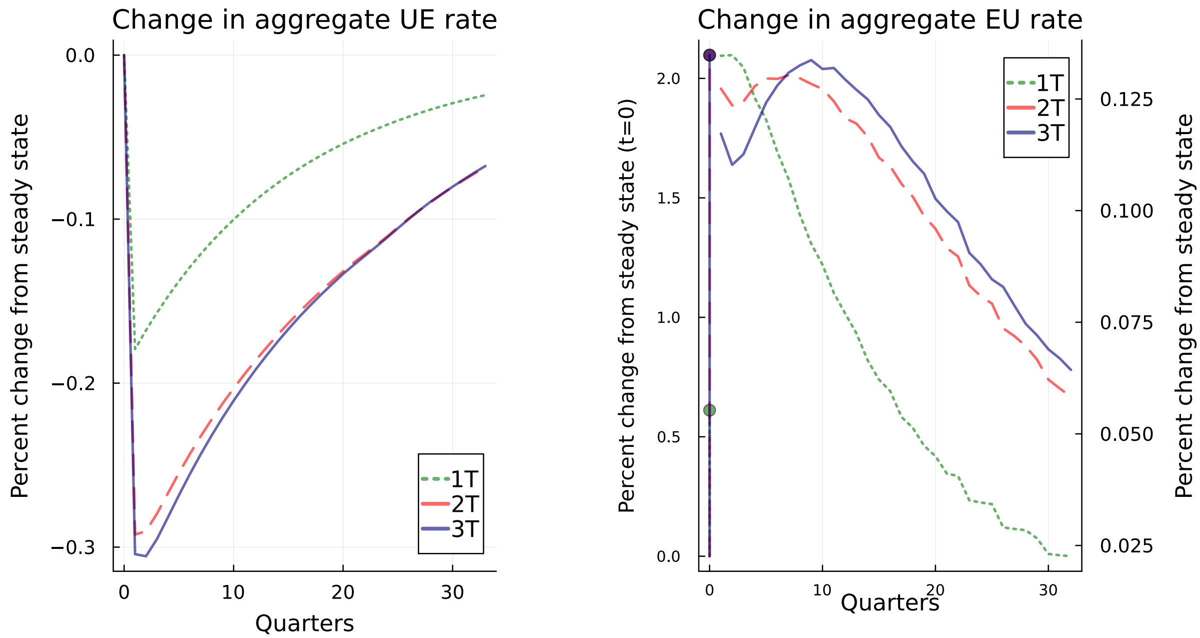

Figure 8: Left panel: model-implied aggregate UE rates for 1T, 2T, 3T models; right panel: model-implied aggregate EU rates for 1T, 2T, 3T models (left axis refers to the response on impact, right axis refers to everything thereafter)

by 10% and recovers very slowly. The difference in the magnitude of the response of the EU rate is due to the fact that the density of the cross-sectional distribution of employed workers around the reservation quality is lower in a representative-worker model than in a model with heterogeneous workers. The difference in the persistence of the response of the EU rate is due to the fact that the search process of the representative displaced worker looks more like the search process of an α-worker (i.e. fast) than of a γ-worker (i.e. slow and with multiple unemployment spells). In turn, this is because αs are the majority of workers and, hence, their behavior dominates the calibration of the representative-worker model.

In the 1T model, the shock leads to a 3 percentage points increase in the unemployment rate with a half-life of 3 years. In the 3T model, the shock leads to a 5.5 percentage point increase in the unemployment with a half-life of 5 years. These differences result from the differences in the magnitude and persistence of the response of the UE and EU rates. In the 1T model, the shock leads to a 9% decline in labor productivity with a half-life of 3 years. In the 3T model, the shock leads to a 7.5% decline in labor productivity with a half-life of 2 years. Intuitively, in the representative-worker model, the shock does not cleanse the employment pool from low-productivity workers and, hence, labor productivity falls by more. Moreover, in the representative-worker model, the composition of the employment pool remains constant over time and, hence, labor productivity recovers at the same speed as the aggregate component of productivity.

Overall, the response of the labor market to an aggregate productivity shock is very different if one completely abstracts from workers' heterogeneity. In particular, completely abstracting from workers' heterogeneity leads to underestimate the magnitude of the increase in the unemployment rate by 2.5 percentage points (45%), underestimate the half-life of the increase in unemployment by 2 years (40%), overestimate the magnitude of the decline in labor productivity by 1.5 percentage points (20%) and overestimate the persistence of the decline in labor productivity by 1 year (30%).

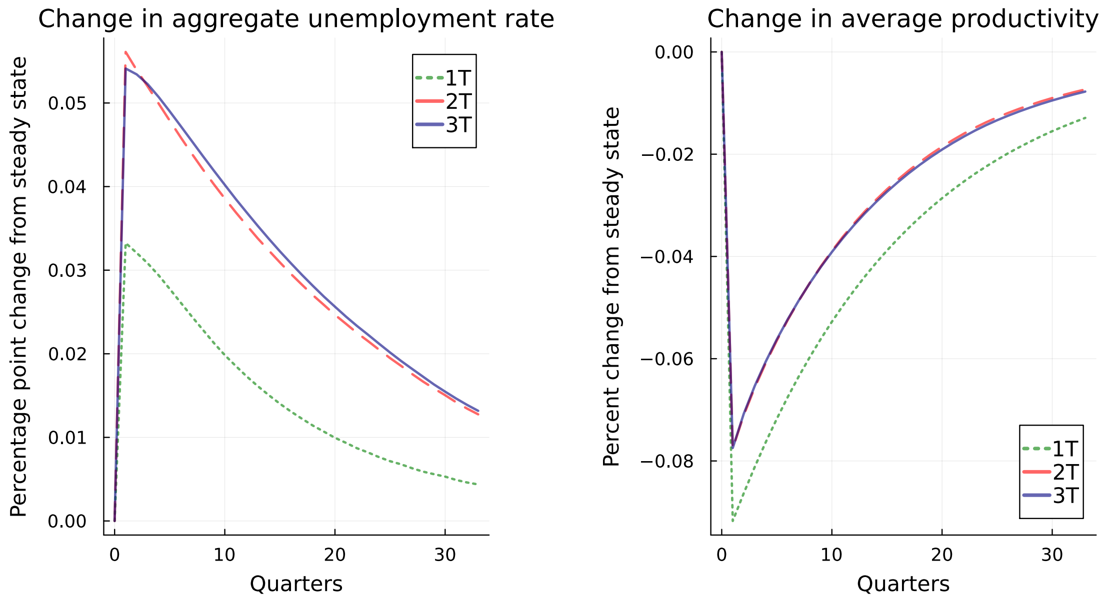

Figure 9: Left panel: response of the aggregate unemployment rate for 1T, 2T, 3T models; right panel: response of the aggregate productivity of labor for 1T, 2T, 3T models

In Figures 8 and 9, we also report the response of the 2T model to the negative shock to aggregate productivity. The responses of the aggregate UE and EU rates are nearly identical in the 2T and in the 3T models. Similarly, the responses of aggregate unemployment and average labor productivity are nearly identical in the 2T and in the 3T models. These findings should not be surprising. As discussed in the previous section, the response of the labor market to an aggregate productivity shock is mainly driven by the behavior of γs. The 2T model accounts for the difference between γs and other workers and, for this reason, it successfully reproduces the labor market response of the baseline model.

We are not the first to point out that workers' heterogeneity is important to understand aggregate labor market fluctuations, but we are the first to measure the extent of workers' heterogeneity and to quantify its impact on the cyclical behavior of the labor market. Pries (2007) considers a model in which there are two types of workers, who differ with respect to their productivity, the average of their exogenous EU rate, and the elasticity of their exogenous EU rate with respect to aggregate productivity shocks. Pries (2007) finds that—if the EU rate of low-productivity workers is higher and more elastic than the EU rate of high-productivity workers—the volatility of the aggregate unemployment rate increases substantially compared to a representative-worker model. Intuitively, the fact that low-productivity workers have a higher EU rate implies that they represent a disproportionate fraction of the unemployed. And the fact that low-productivity workers have a more elastic EU rate implies that their share in the unemployment pool is countercyclical. The two effects combine to generate average gains from trade that are lower and more countercyclical than in a representative-worker model. In turn, this leads to an unemployment rate that is more volatile than in a representative-worker model.

Ferraro (2018) considers a model where workers are heterogeneous with respect to their individual productivity and populate segmented sections of the labor market. When the economy is hit by a negative shock to aggregate productivity, the response of the UE rate is larger for lower productivity workers, as they have smaller gains from trade. Moreover, when the economy is hit by a negative shock to aggregate productivity, the least productive workers find it optimal to move from employment into unemployment. The extreme volatility of unemployment for the least productive workers generates large fluctuations in aggregate unemployment. The low volatility of unemployment for the most productive workers keeps fluctuations in average labor productivity small.

While the mechanism for amplification in Pries (2007) and Ferraro (2018) is similar to ours, these papers are more of a "proof of concept", since they do not offer a credible quantification of the mechanism.[22] In Pries (2007), the exogenous link between a worker's type, his productivity, and his pattern of employment transitions is not based on empirical evidence—since doing so would require estimating fixed-effects, an exercise that Pries acknowledges would require longitudinal data. Similarly, in Ferraro (2018), the endogenous link between a worker's type, his productivity, and his pattern of employment transitions is not validated at all.[23] In contrast, we directly estimate workers' heterogeneity from fixed-effects in a large longitudinal dataset, we provide a theoretical interpretation of heterogeneity, and we validate our interpretation of heterogeneity at the micro level. Moreover, in the next section, we bring direct evidence of the role played by different types of workers in shaping the evolution of the aggregate labor market during and after the Great Recession, and show that this evidence agrees with the macroeconomic predictions of the theoretical model.

6.3 Worker types in the Great Recession

We now want to use our panel dataset to document the dynamics of the labor market during and after the Great Recession of 2008-2009. We compare the actual behavior of aggregate unemployment and type-specific unemployment rates with those implied by our theory in response to an aggregate productivity shock. We find that, even though recessions are likely caused by a multiplicity of shocks (productivity shocks, demand shocks, policy shocks, etc…), the labor market dynamics predicted by the theory align quite well with the dynamics of the actual labor market. As predicted by the theory, the increase in the unemployment rate was much smaller for αs than for γs, the re-absorption of the excess unemployment rate was much faster for αs than for γs, and, as a result, the slow recovery of the labor market that characterized the aftermath of the Great Recession was entirely driven by the behavior of gamma workers. These findings are reassuring, as they provide some indirect validation to our theoretical interpretation of workers' heterogeneity. Specifically, our theory interprets the low earnings of γs

[22]It is important to stress that the models of Pries (2007) and Ferraro (2018) do not have the propagation mechanism that is a key part of our model. In our model, aggregate excess unemployment is re-absorbed slowly because displaced γs experience multiple unemployment spells before finding another stable job, thus keeping their EU rate and their unemployment rate elevated long after an aggregate shock has hit the economy. This propagation mechanism is based on the observation that γs rarely keep a job for more than 2 years, an observation that is rationalized by having γs sample from a match-quality distribution with a thick right tail. Neither Pries (2007) nor Ferraro (2018) allow for match-quality heterogeneity and, hence, neither of them captures our mechanism. Moreover, the propagation mechanism is quantitatively important. When we calibrate a version of our model that abstracts from match-quality heterogeneity, we find that the response of unemployment to an aggregate shock is about 40% less persistent.

[23]The relationship between types, productivity and transitions conjectured by Pries (2007) and Ferraro (2018) is not even entirely consistent with the data at a qualitative level. Pries (2007) conjectures that low-productivity types have a higher and more cyclical EU rate (which we find to be true) and the same UE rate as high-productivity types (which we find to be false). In Ferraro (2018), all types of workers have the same, constant exit rate from a job, unless a negative aggregate shock makes their type-specific gains from trade negative—in which case their exit rate is 1 and their UE rate is approximately zero (not exactly zero because of an artifact of modelling the economy in discrete time). In contrast, we find that different types have a systematically different, tenure-dependent exit rate from a job even before the Great Recession.

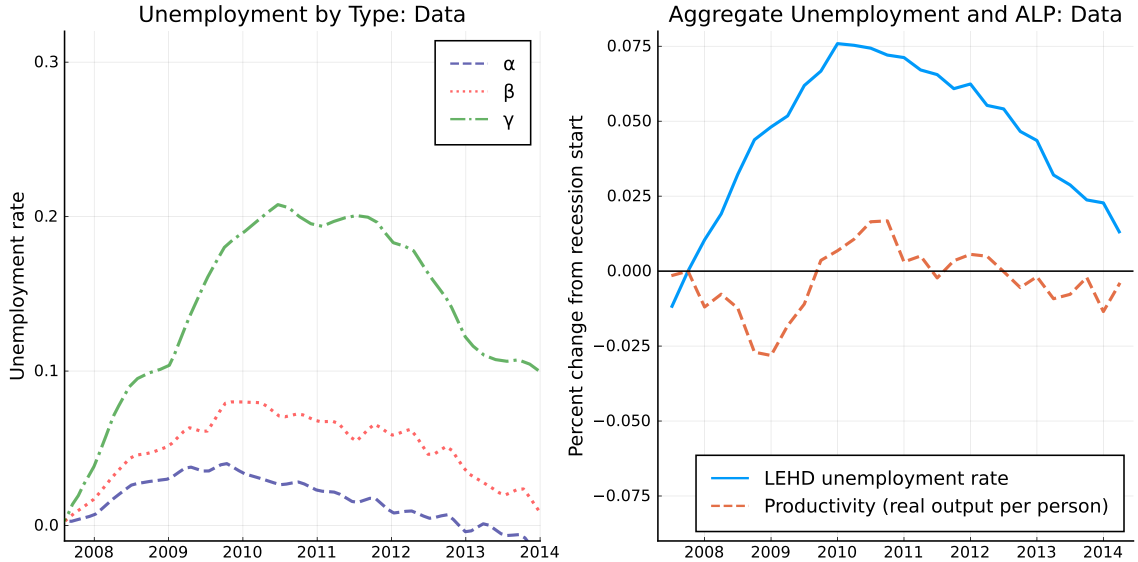

Figure 10: Left panel: increase in unemployment rates for each worker type during the Great Recession in the data; right panel: LEHD unemployment and average labor productivity during the Great Recession in the data

as evidence of small gains from trade in the labor market, and the distribution of job durations of γs as evidence of a match-quality distribution with a decreasing density and a long right tail. The theoretical interpretation of the evidence implies, in turn, that aggregate productivity shocks would have a large and persistent impact on the unemployment rate of γ workers. Finding empirical confirmation of these implications provides indirect validation to our theoretical interpretation of the data.[24]

The left panel of Figure 10 plots the unemployment rate of different types of workers during and after the Great Recession. 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.[25] The unemployment rate of αs increases by 3 percentage points over the period 2008-2010, and quickly falls down afterwards. By the end of 2012, the unemployment rate of 3b1s

[24]Some of our readers may be concerned that γ-workers are simply the workers who happened to be most affected by the Great Recessions. While this is a reasonable concern, there are two reasons why we are not worried about it. First, while we classify workers based on their entire biography (including the Great Recession), we calibrate the model to type-specific moments computed only for the pre-Great Recession period. If γs were simply the workers that happened to be hit the hardest by the recession, one would expect the pre-recession moments that describe γs to be similar to the pre-recession moments that describe αs. This is not the case. The difference between the moments that describe αs and γs over the entire period of observation and over the pre-recession period are quite similar. Second, when we compare the classification of workers based on the entire period of observation with a classification of workers based only on observations before the recession, we find a high degree of correlation. This suggests that the workers that we identify as γ in the paper would be also classified as γ using only pre-recession data.

[25]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.

is back to its pre-recession level. The unemployment rate of βs 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 γs increases by a staggering 21 percentage points between 2008 and 2010 and, afterwards, it falls back very slowly. The right panel of Figure 10 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 3 percent from trend from 2008 to 2009 and then recovers quickly, returning to trend by 2010.

The picture in the left panel of Figure 10 is dramatic, and provides direct evidence of the role played by workers' heterogeneity in shaping labor market fluctuations. The increase in the unemployment rate of γs is 3 times as large as the increase in the unemployment rate of βs, and about 7 times as large as the increase in the unemployment rate of αs. Moreover, note how excess aggregate unemployment would have been re-absorbed quickly had it not been for γs—whose excess unemployment was still hovering at about 10 percentage points in 2014.

The picture in the left panel of Figure 10 is similar to the one in the left panel of Figure 6—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 γs, by 5 percentage points for βs, and by 2.5 percentage points for αs. That is, the model predicts that the productivity shock generates responses in the unemployment rate for different types that are of the same order of magnitude as what we observe in the Great Recession. Second, the model predicts that the productivity shock generates responses in the unemployment rate that have different persistence for different types—most persistent for γs and least persistent for αs. This is precisely what we observe in the aftermath of the Great Recession. Third, the model predicts that the productivity shock generates a decline in labor productivity that is less persistent than the increase in unemployment. This is also consistent with what we see in the data.

According to our theory, the unemployment of γs is so persistent because displaced γ workers go through several spells of unemployment before finding a stable match in the right tail of the quality distribution. The left panels of Figure 11 plot the cumulated number of completed unemployment spells in the data for αs, βs and γs who are displaced at the beginning of the recession and for αs, βs and γs who are not. As expected, the cumulated number of completed unemployment spells is higher for displaced γs (4 spells), than for βs (3 spells) and αs (2 spells), and much higher than for non-displaced workers (about 1 spell for all types). The right panels of Figure 11 plot the cumulated number of completed spells in the Great Recession for αs, βs and γs who are and are not displaced in the model. The figures are very similar.

The model, however, appears to miss the exact 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 3%. A natural and simple alternative to a shock to the aggregate component of productivity x are shocks to the type-specific components of productivity y. One interpretation of y-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-shocks is that other shocks, say financial or demand shocks, impact the value of their output differently. Leaving aside issues of interpretation, we simply want to understand whether we can find a series of y-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

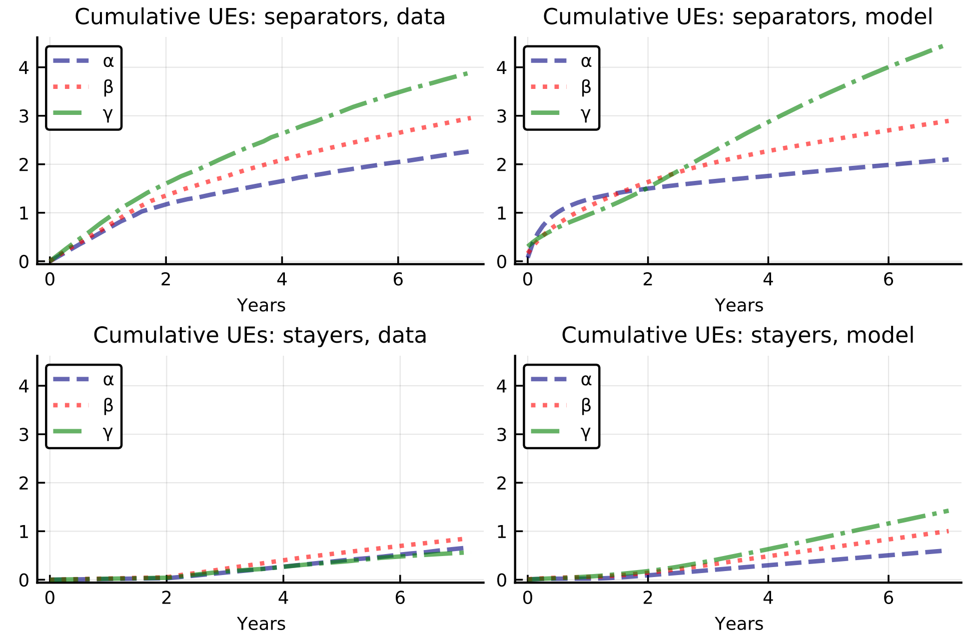

Figure 11: Cumulated number of unemployment spells/UE transitions for those who are displaced within one year of the start of the recession and those who are not.

the Great Recession.

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 αs than for βs and γs. Specifically, at its minimum, the type-specific productivity is 5% lower than in steady state for αs, and 12% lower than in the steady state for βs and γs.

The left panel of Figure 12 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 αs, by 6.5 percentage points for βs, and by 23 percentage points for γs. 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 αs is fully reabsorbed in 2012. The increase in the unemployment rate of βs is fully reabsorbed in 2013. For γs, 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 recession. Both the magnitude and the persistence of the increase in aggregate unemployment are driven by the response of the unemployment rate for γs. 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

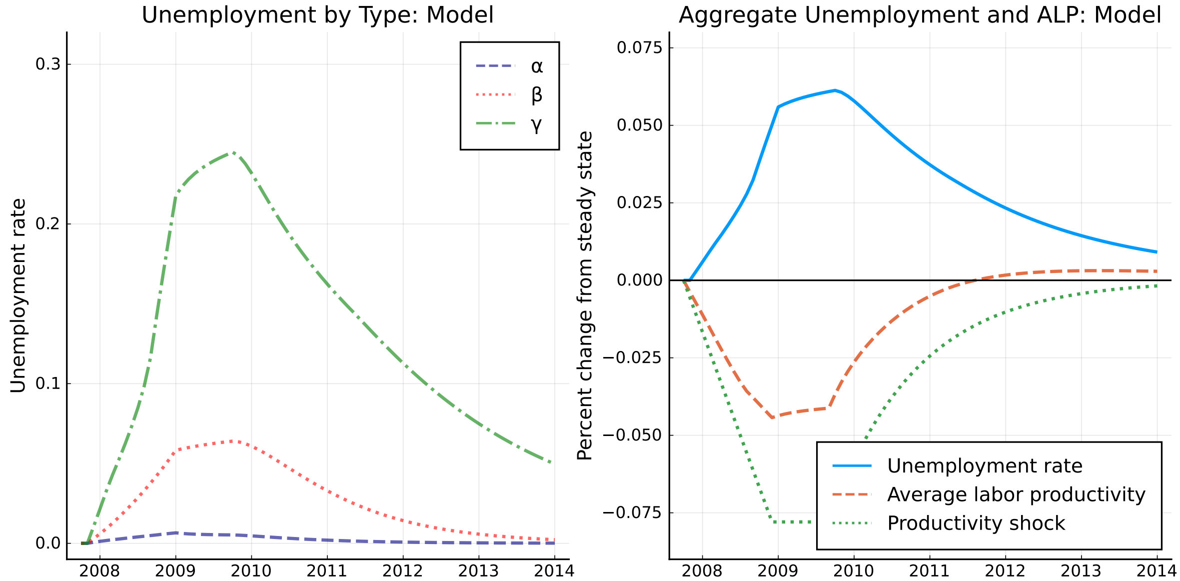

Figure 12: Left panel: increase in unemployment rates for each worker type during the Great Recession in the model; right panel: LEHD unemployment and average labor productivity during the Great Recession in the model

a strong cleansing effect across types. Indeed, since the decline in productivity is larger for βs and γs, the composition of the employment pool tilts strongly towards αs 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 αs.

A comparison between Figures 10 and 12 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 very well the behavior of the US labor market during and after the Great Recession.[26] 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 (γs) and the types that drive the dynamics of labor productivity (αs).

7 Conclusions

We accessed a long and large panel dataset of US workers to measure the extent to which individuals differ with respect to their pattern of employment transitions. We used the data to calibrate, via the two-stage Grouped-Fixed Effects method of Bonhomme, Lamadon and Manresa (2021), a search-theoretic model of the labor market in the style of Menzio and Shi (2011), in which workers are heterogeneous with respect to the parameters that control their stochastic process of transitions across employment states. We found that heterogeneity can be discretized with three types: αs, βs and γs. Workers of type α have a high UE rate because

[26]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).

they have large gains from trade, and they stick to their jobs because they do not face much heterogeneity in match quality. Workers of type γ have a low UE rate because they have smaller gains from trade, and are unlikely to stick to their job because they keep searching for the right tail of the match-quality distribution. We used the calibrated model to measure the impact of an aggregate productivity shock on the labor market. We found that aggregate unemployment is elastic and persistent because the unemployment of γs is elastic and persistent. In line with the predictions of the theory, we documented that the unemployment rate in the Great Recession increased so much and recovered so slowly because of γs.[27]

Much work remains to be done. First, we need to understand what determines the type of a worker, since basic demographics, industry and location cannot forecast a worker's type. In order to shed more light on this question, we would need to access a dataset containing more information about individuals (such as, say, the NLSY). Second, we need to understand whether a worker's type is permanent or transitory. Given the limitations of our data, we maintained that a worker's type is fixed—even though a preliminary analysis seems to show that workers' types are highly correlated over time. Relatedly, it would be interesting to access a longer panel dataset to understand whether the distribution of types in the workforce has changed over time and, in turn, whether such change may be responsible for the changing nature of labor market fluctuations. Lastly, it would be interesting to connect the insights of this paper to optimal policy. For instance, it would be interesting to understand how we should redesign unemployment insurance given that workers are heterogeneous with respect to the nature of their job search.

[27]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. 2020. “Heterogeneity and Unemployment Dynamics” *Journal of Business & Economic Statistics.*

[3] Ahn, H., B. Hobijn, and A. Şahin. 2023. “The Dual US Labor Market Uncovered.” NBER Working Paper 31241.

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

[5] 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.

[6] Bonhomme, S., T. Lamadon, and E. Manresa. 2021. “Discretizing Unobserved Heterogeneity,” *Econometrica.*

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

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

[9] Darougheh, S., and G. Lundgren. 2022. “Worker Protection and Screening.” Manuscript, Denmark Central Bank.

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

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

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

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

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

[15] Fukui, M. 2021. “A Theory of Wage Rigidity and Unemployment Fluctuations with On-the-Job Search,” Manuscript, MIT.

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

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

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

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

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

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

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

[23] 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.

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

[25] Honore, B. 1993. “Identification Results for Duration Models with Multiple Spells.” *Review of Economic Studies*, 60: 241-246.

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

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

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

[29] 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*.

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

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

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

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

[34] Martellini, P., G. Menzio, and L. Visschers. 2021. “Revisiting the Hypothesis of High Discounts and High Unemployment.” *Economic Journal*, 131: 2203-2232.

[35] Menzio, G. 2022. "Stubborn Beliefs in Search Equilibrium." *NBER Macro Annual*, 2022, 37, 239-297, edited by M. Eichenbaum, E. Hurst and V. Ramey, University of Chicago Press.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[50] Shibata, I. 2015. "Labor Market Dynamics: A Hidden Markov Approach." Manuscript, University of Chicago.

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

[52] Spinella, S. 2022. "Discretizing Worker Heterogeneity: The Alpha Beta Gamma of the Italian Labor Market." Manuscript, Università Bocconi.

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