Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

The Role of Observed and Unobserved Heterogeneity in the Duration of Unemployment Spells

Hie Joo Ahn

2016-063

Please cite this paper as:
Ahn, Hie Joo (2022). "The Role of Observed and Unobserved Heterogeneity in the Duration of Unemployment Spells," Finance and Economics Discussion Series 2016-063r1. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2016.063r1.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

The Role of Observed and Unobserved Heterogeneity in the Duration of Unemployment Spells

Hie Joo Ahn [‡]

Federal Reserve Board

March 1, 2022

Abstract

This paper studies the degree to which observable and unobservable worker characteristics account for the variation in the aggregate duration of unemployment. I model the distribution of unobserved worker heterogeneity as time varying to capture the interaction of latent attributes with changes in labor-market conditions. Unobserved heterogeneity is the main explanation for the duration dependence of unemployment hazards. Both cyclical and low-frequency variations in the mean duration of unemployment are mainly driven by one subgroup: workers who, for unobserved reasons, stay unemployed for a long time. In contrast, changes in the composition of observable characteristics of workers have negligible effects.

Keywords: unemployment duration, disaggregate unemployment, unobserved heterogeneity, genuine duration dependence, nonlinear state space model, extended Kalman filter

[*]The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. I thank Gianni Amisano, Travis Berge, Michael Boerman, Marco Del Negro, James Hamilton, Andrea Stella, Caitlin Hesser, David Jenkins and Christopher Karlsten for helpful comments on an earlier draft of this paper.

[†]E-mail: hiejoo.ahn@frb.gov

Introduction

Why does the average duration of unemployment rise during an economic downturn? Darby et al. (1986) argue that recessions are periods when workers with low reemployment prospects lose their jobs, and the increased share of these workers in the unemployment pool raises the mean duration of unemployment. Baker (1992) labels this explanation the "heterogeneity hypothesis." A quite different view is that during a recession, the individual duration of unemployment becomes longer due to weak labor demand, regardless of the worker's characteristics.

This paper revisits this debate—a debate that has been at the center of the discussion on labor market dynamics for decades—using an empirical approach with two novel aspects. First, my approach takes into account worker characteristics that are both observed and unobserved in the data, while previous studies (e.g., Kroft et al. (2016)) only consider observed attributes. Hereafter, I will refer to observed and unobserved worker characteristics as observed and unobserved heterogeneity, respectively. Second, I model the distribution of unobserved heterogeneity to be time varying, whereas the existing literature only characterizes it as time invariant (e.g., Shimer et al. (2015)).

The empirical findings in this paper support the *"unobserved" heterogeneity hypothesis*. We postulate that there is a group of workers who are prone to long-term unemployment due to their unobserved attributes. Such workers make up a larger fraction of those who have been unemployed for a long period of time compared to a group of workers who have been unemployed for a short period of time. This could account for the negative duration dependence of unemployment hazards, which is the observation that the unemployment-exit probability declines with unemployment duration. When there is a negative economic shock, disproportionately more workers in this subgroup flow into the unemployment pool than others. As a result, the exit from unemployment as well as the entry into unemployment of this subgroup of workers could be the main driver of cyclical and low-frequency variations in the mean duration of unemployment. The empirical conclusion of this paper is that this is indeed what happens.[1]

Why do we need to consider a time-varying distribution for unobserved heterogeneity? Previous studies treat unobserved heterogeneity as a fixed characteristic with constant effects on the individual's job-finding prospects over time. However, this assumption is not valid if unobserved worker characteristics interact with changes in labor-market conditions. Consider for example an economy where workers with certain skills become less productive and thus are less in demand because of skill-biased technological changes (Acemoglu (2002)). Workers with those skills tend to have lower job-finding probabilities and experience longer unemployment spells, fattening the right tail of the unemployment duration distribution. During an economic downturn, firms might shed these workers as they become unable to retain less productive workers. The job-finding probabilities of such workers fall further. Therefore, their share in unemployment rises, driving up long-term unemployment. In such an economy, the mean duration of unemployment goes up during the recession because workers with skills that are in lower demand make up a larger fraction of the pool of unemployed workers. If the skill differences are not well captured by observable characteristics, we would observe a common increase in unemployment duration across workers with the same observable characteristics, incorrectly concluding that worker heterogeneity is unimportant. The role of worker heterogeneity in the aggregate duration of unemployment will not be accurately measured, if we fail to consider the time-varying distribution of unobserved heterogeneity.

[1] Previous studies such as Krueger et al. (2014) and Kroft et al. (2016) claim that aggregate factors that affect different workers in a similar way, rather than changes in the composition of the unemployed, are the cause of the record-high level of unemployment duration after the end of the Great Recession. However, these studies do not consider unobserved worker heterogeneity. As further support for the conclusions of this paper, Gregory et al. (2021) find that a small group of workers with a particular unobserved attribute—what they call "γ-workers"—mainly accounts for the sharp rise in unemployment during the Great Recession and its slow recovery. For the identification of unobserved types, the authors apply k-means clustering to the Longitudinal Employer-Household Dynamics dataset, while I estimate a nonlinear state space model with the dataset constructed from the micro data of Current Population Survey. Gregory et al. (2021) provide a search-theoretic model that accounts for the empirical finding, while this paper does not.

How do we identify workers with unobserved attributes? Suppose unemployed individuals are one of two unobserved types who have ***h***igh (H) and ***l***ow (L) probabilities of exiting unemployment status next month. Newly unemployed individuals of both types flow into the unemployment pool each month. I will refer to newly unemployed individuals as the *inflows*. Suppose that the economy is in a steady state. Using the inflows and the probabilities of continuing unemployment next month for the two types, we can calculate the number of individuals unemployed for n months. This means that given four observations over the full sample—the average number of individuals unemployed for 1 month, 2 to 3 months, 4 to 6 months, and 7 to 12 months—we can identify the four population parameters—inflows and unemployment-continuation probabilities of types H and L. More generally, suppose that a worker's chance of exiting unemployment changes with the length of the worker's unemployment spell, which is often referred to as genuine duration dependence (hereafter, GDD). For example, employers might discriminate against the long-term unemployed, or workers might take a low-paying job as their jobless spells get longer. If GDD is a linear function of the unemployment duration characterized by one parameter, we can further identify the GDD parameter together with an additional data point—the number of individuals unemployed for longer than one year. In fact, we observe these five data points every month from January 1976 to December 2019. Ahn and Hamilton (2020) show that with such data, we can identify time-varying inflows and unemployment-continuation probabilities of two unobserved types at each point in time, while allowing a more flexible functional form for GDD.

I apply the identification scheme of Ahn and Hamilton (2020) to unemployment data disaggregated on the basis of observable characteristics to estimate the inflows and unemployment-continuation probabilities of workers of unobserved types who share the same observable characteristics. With the estimates, I recover the distribution of unemployment duration for each worker type at each point in time.[2] I find substantial heterogeneity in unemployment hazards even within a group of unemployed individuals who share the same reported reason for unemployment, which is the key observable characteristic that seems to account for differences in job-finding rates of unemployed individuals (Fujita and Moscarini (2017) and Hall and Schulhofer-Wohl (2018)). I also find unobserved heterogeneity is crucial among a group of workers with the same demographic characteristics or level of education. This result implies that the long-term unemployed mostly represent workers with some unobserved attributes, and their inflows and unemployment-continuation probability drive the variation in long-term unemployment.

Unobserved heterogeneity is crucial in understanding changes in the aggregate duration of unemployment. Using a shift-share analysis, I show that from December 2007 to December 2011—the period during which the mean duration of unemployment registers its most dramatic increase—the rise in unemployment duration of type L workers is responsible for about 80% of the observed hike in the mean duration. The remaining 20% is explained by the increased share of type L workers in the unemployment pool. The compositional shift of workers with observable characteristics is essentially irrelevant in this period. It is further notable that during the Great Recession, aggregate unemployment duration rises mainly due to the increased share of type L workers. During the recovery, however, the share of type L workers starts to slowly decline, while their unemployment duration continues to rise and drives up the aggregate duration of unemployment.

Type L workers are also important in secular changes in the aggregate duration of unemployment. Between January 1980 and December 2019, the mean duration of unemployment doubled from 10.4 weeks to 20.8 weeks.[3] The compositional shift of type L workers explains

[2] This paper is closely related to Ahn and Hamilton (2020). The key differences are that while Ahn and Hamilton (2020) focus on cyclical unemployment variations and do not consider the observable characteristics of workers, this paper concentrates on the determinants of unemployment duration and explicitly considers the observable characteristics as potential drivers. Ahn and Hamilton (2020) cite this paper in framing their discussion on the association between reasons for unemployment and unobserved types.

[3] The difference between the two levels is likely to represent the trend, as the two dates represent business-

30% of the secular rise and type L duration explains 65%. Both factors prevent the mean duration of unemployment from recovering to pre-recession levels after each recession ends, thereby driving the uptrend in the average duration of unemployment.

This paper is composed of five sections. Section 1 discusses the data used for the empirical analysis. Section 2 illustrates the need to consider unobserved heterogeneity in understanding the distribution of unemployment duration. Section 3 introduces the empirical methodology, and Section 4 discusses the empirical results. Section 5 analyzes the contribution of worker heterogeneity to the evolution of aggregate unemployment duration. Section 6 explores the source of low-frequency variation in the mean duration of unemployment.

1 Data

The empirical exercise is based on the numbers of people who have observed characteristic j and have been unemployed for 1 month (less than 5 weeks), 2-3 months (5-14 weeks), 4-6 months (15-26 weeks), 7-12 months (27-52 weeks), and longer than 1 year (53 weeks and over) in each month t.[4] I denote these five numbers by U_{jt}^{1}, $U_{jt}^{2.3}$, $U_{jt}^{4.6}$, $U_{jt}^{7.12}$, and $U_{jt}^{13.+}$, respectively. Each value of j summarizes the observable characteristics of unemployed individuals including demographic characteristics, education level, previous industry and occupation, and reason for unemployment. In particular, I consider five reasons for unemployment: temporary layoffs, permanent separation, job leavers, reentrants to the labor force, and new entrants to the labor force.[5] For the aggregate data, the notation j is suppressed. I construct

cycle peaks.

[4]The Bureau of Labor Statistics (BLS) reports the number unemployed for less than 5 weeks, 5-14 weeks, 15-26 weeks, and 27 weeks and over to minimize measurement errors (e.g., digit preference) by averaging within broad duration groups. Within long-term unemployment, the BLS often further breaks down the number unemployed for 27 weeks and over into those unemployed for 27-52 weeks and those with a duration longer than 52 weeks. Because the main focus of this paper is how unemployed individuals' labor force transitions between months affect the distribution of unemployment duration, I use "months" as the unit of unemployment duration instead of "weeks."

[5]The Current Population Survey (CPS) asks unemployed individuals in which circumstance they become unemployed. There are five reasons for unemployment in the CPS that are temporary layoffs, permanent job loss, job leavers, reentrants to the labor force, and new entrants to the labor force. Permanent job loss can be

the dataset using the CPS micro data.[6] The sample period is January 1976–December 2019.

2 Why study heterogeneity within an observed category?

Why is it important to consider heterogeneity within an observed category when analyzing the duration of unemployment? This section illustrates that a model that does not consider unobserved heterogeneity or GDD is limited in its ability to predict the distribution of unemployment duration in the data.

Consider an economy in which unemployed individuals have the same probability of exiting unemployment at t conditional on being unemployed at $t-1$. Let U_t denote the total number of unemployed individuals and U_t^1 denote the number of newly unemployed individuals with the duration of unemployment one month in month t. The probability of continuing to be unemployed at t conditional on being unemployed at $t-1$, p_t, is calculated from

$$p_t = \frac{U_t - U_t^1}{U_{t-1}}.$$

In this economy, the number of those unemployed for n months at t, denoted $\hat{U}_t^n$, is determined by how many people become newly unemployed at $t-n+1$ and by the history between $t-n+1$ and t of the probability of staying unemployed next month conditional on being unemployed in the current month. Then, $\hat{U}_t^n$ is written into the following:

$$\begin{aligned} \hat{U}_t^1 &= U_t^1 \\ \hat{U}_t^n &= U_{t-n+1}^1 \prod_{h=2}^{n} p_{t-n+h} \qquad \text{for} \quad n \geq 2. \end{aligned} \tag{1}$$

Suppose that the maximum duration of unemployment is 48 months.[7] With $\hat{U}_t^n$ with $n =$

further divided into temporary job ended and other separation, but this categorization is available after 1994. Therefore, I use the five-way breakdown for the empirical exercise, as the sample period is 1976-2019.

[6]Further details on the data construction are found in the online appendix.

[7]In the CPS, the maximum duration of unemployment was 2 years before January 2011 but was extended to

$1, 2, ..., 48$, we can predict the mean and standard deviation of unemployment duration in progress in each month in this economy. Let M_t^{agg} denote the mean, and S_t^{agg} denote the standard deviation.

$$
\begin{aligned}
M_t^{agg} &= \frac{\sum_{n=1}^{48} \hat{U}_t^n n}{\sum_{n=1}^{48} \hat{U}_t^n} \\
S_t^{agg} &= \sqrt{\frac{\sum_{n=1}^{48} \hat{U}_t^n (n - M_t^{agg})^2}{\sum_{n=1}^{48} \hat{U}_t^n}}
\end{aligned} \tag{2}
$$

In Figure 1, Panel A shows M_t^{agg} (red line) and the actual mean duration computed from the CPS micro data (blue line); Panel B plots S_t^{agg} (red line) and the actual standard deviation of unemployment duration (blue line) from January 1980 to December 2019.[8] On average, the predicted mean is 65% of the actual mean, and the predicted standard deviation is 40% of the actual standard deviation. This result suggests that the distribution of unemployment duration and its variation over time cannot be correctly described without taking heterogeneity in unemployment hazards into account.

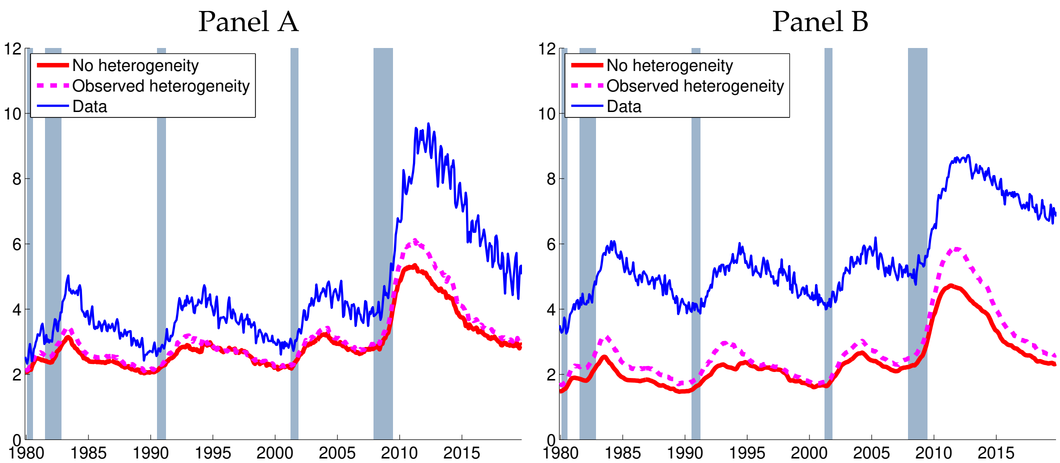
Now suppose that workers differ only by observable characteristics. I will refer to individuals who have observable characteristic j as group j for $j = 1, 2, 3, ..., J$ with J being the number of worker characteristics. Let p_{jt} denote the unemployment-continuation probability at t of individuals in group j. The probability, p_{jt}, is calculated as

$$
p_{jt} = \frac{U_{jt} - U_{jt}^1}{U_{j,t-1}},
$$

where U_{jt} is the total number of unemployed individuals and U_{jt}^1 is the number of newly

5 years from 2011. Before 2011, any response of unemployment duration greater than 2 years was entered as 2 years. In spite of the increased upper bound from 2011, the number of individuals reporting a duration longer than 4 years is low. Therefore, setting the upper bound to be 4 years is not restrictive. Assuming 2 or 3 years as the maximum duration does not change the main result of this paper. Hornstein (2012) also assumes that the maximum duration of unemployment is 4 years in his model of unemployment accounting identity.

[8] Originally, the sample period of micro data is from January 1976 to December 2019. The data from the first four years are used to compute the full distribution of unemployment duration in progress for the first month in the sample, January 1980.



Note: Units are in months. The actual mean and standard deviation are computed from the CPS micro data, and are not seasonally adjusted. Shaded areas denote NBER recessions.
Source: Author's calculation based upon the CPS micro data.

Figure 1: Mean (Panel A) and standard deviation (Panel B) of the distribution of unemployment duration in progress predicted from equations (2) and (3)

unemployed individuals in group j at t. Similar to equation (1), the number of individuals who have been unemployed for n months in group j, denoted by $\hat{U}^n_{jt}$, is written into the following:

$$\begin{aligned} \hat{U}^1_{jt} &= U^1_{jt} \\ \hat{U}^n_{jt} &= U^1_{j,t-n+1} \prod_{h=2}^{n} p_{j,t-n+h} \qquad \text{for} \quad n \geq 2. \end{aligned}$$

The mean duration of unemployment in progress of group j is calculated from

$$M_{jt} = \frac{\sum_{n=1}^{48} \hat{U}^n_{jt} n}{\sum_{n=1}^{48} \hat{U}^n_{jt}}.$$

In this economy, the aggregate mean and standard deviation of unemployment duration, denoted by M_t^{disagg} and S_t^{disagg}, respectively, are calculated as follows:

$$\begin{aligned} M_t^{disagg} &= \sum_{j=1}^{J} (\frac{U_{jt}}{\sum_{j=1}^{J} U_{jt}}) M_{jt} \\ S_t^{disagg} &= \sqrt{\sum_{j=1}^{J} (\frac{U_{jt}}{\sum_{j=1}^{J} U_{jt}}) \frac{\sum_{n=1}^{48} \hat{U}_{jt}^{n} (n - M_{jt})^2}{\sum_{n=1}^{48} \hat{U}_{jt}^{n}}}. \end{aligned} \tag{3}$$

For illustrative purposes, I take the reason for unemployment as the observed category. As shown by Figure 1, M_t^{disagg} (dotted fuchsia line in Panel A) is 70% of the actual mean, and S_t^{disagg} (dotted fuchsia line in Panel B) is 50% of the actual standard deviation. The predicted mean and standard deviation are slightly larger than those simulated under the assumption that all workers are homogeneous.[9] Although we consider different observed characteristics in varying levels of detail, meaningful improvement is not achieved in fitting the distribution of unemployment duration in progress to what is observed in the data.

In the remainder of the section, I argue that it is crucial to consider the duration dependence in unemployment hazards to match the observable distribution of unemployment duration among individuals who share the same detailed observable characteristics. Consider an economy in which unemployed individuals who share the same observable characteristics have different unemployment-continuation probabilities. I postulate that the unemployment-continuation probability of workers in group j is a function of duration, τ, to characterize changes in unemployment hazards over the duration of unemployment. Suppose that the probability, denoted by $p_{jt}(\tau)$, is written into the following form:

$$p_{jt}(\tau) = \exp(-\exp(d_{jt}^{\tau})), \tag{4}$$

[9]By Jensen's inequality, the mean and standard deviation become equal or larger, as we consider finer gradations of heterogeneity in the model.

where d_{jt}^{τ} is a cubic function of τ,

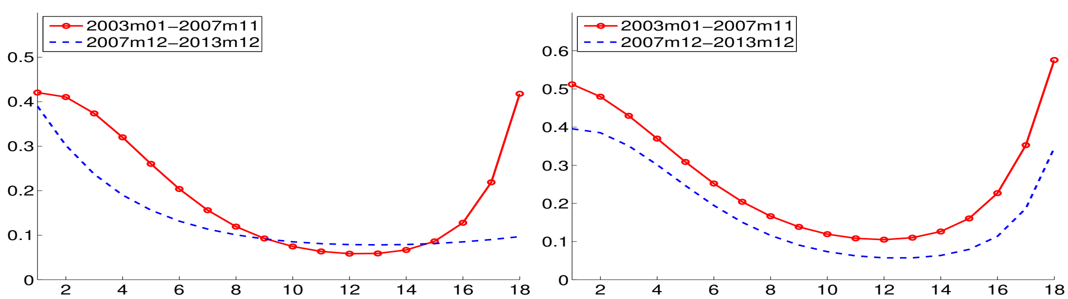
$$d_{jt}^{\tau} = \delta_{jt}^{a} + \delta_{jt}^{b}\tau + \delta_{jt}^{c}\tau^2 + \delta_{jt}^{d}\tau^3.\text{[10]} \tag{5}$$

The double-exponential function is a convenient way of implementing a proportional hazard specification to guarantee a well-defined probability between 0 and 1 (e.g., Katz and Meyer (1990)).

Suppose, for simplicity, that the economy is in a steady state. Then, the number of individuals unemployed for n months in group j, $\hat{U}_j^n$, is written as follows:

$$U_j^n = U_j^1 p_j(1)....p_j(n-1). \tag{6}$$

Note that there are four unknown parameters—$\delta_j^a, \delta_j^b, \delta_j^c$ and δ_j^d. With the observed data—U_j^1, $U_j^{2.3}, U_j^{4.6}, U_j^{7.12}$, and $U_j^{13.+}$—we can solve for the four unknown parameters with the four equations for $U_j^{2.3}, U_j^{4.6}, U_j^{7.12}$, and $U_j^{13.+}$.



Men, 25-44, Some college/associate degree

Women, 16-24, High school or less

Source: Author's calculation.

Figure 2: Duration profile of exit probability from unemployment, $1 - p_j(\tau)$. Horizontal axis: duration of unemployment in months. Vertical axis: probability that an individual leaves the unemployment status the following month.

[10]A cubic function can flexibly characterize both linear and nonlinear functions. I adopt a cubic function to capture possible nonlinearity in the duration profile of unemployment hazards for illustrative purposes.

I estimate the model, equations (4)-(6), for two different groups—men aged 25-44 with some college education or an associate degree and women aged 16-24 with a high school education or less—and for two different periods—January 2003-November 2007 and December 2007-December 2013.[11] Figure 2 plots the average exit probability from unemployment over unemployment duration, $1 - p_j(\tau)$, for each group and for each period of time. The duration profile of unemployment hazards is different over business cycle phases and between different groups.

There are three important observations. First, the unemployment-exit probabilities exhibit a non-monotonic U-shape along the duration of unemployment, decreasing for the first year of duration in unemployment and increasing afterward. Deterioration in unemployment hazards over the duration—the negative duration dependence—is commonly observed in the two groups and in the two different time periods during the first year of unemployment. There are two explanations for the negative duration dependence. One theory claims workers with low re-employment prospects due to their unobserved attributes are dynamically sorted into the long-term unemployment pool, creating the observed deterioration in unemployment hazards. Another theory suggests workers' re-employment prospects decline during their jobless spells, because potential employers discriminate against the long-term unemployed or their human capital depreciates over time. This theory is called the *negative* GDD. Though these two theories explain the negative duration dependence, the rising exit probabilities after one year in unemployment suggest that workers are exposed to a different type of GDD. The long-term unemployed might become more likely to leave the labor force out of discouragement or to take a low-paying job due to their

[11] I consider the two groups to illustrate that the duration dependence is a general phenomenon commonly observed across individuals after controlling for demographic and education characteristics. In addition, I focus on the two periods of time with different average unemployment rates (5.2% in 2003:M01-2007:M11 and 8.2% in 2007:M12-2013:M12) to illustrate that the pattern of duration dependence can change depending on the level of the unemployment rate. The nonlinear state space model introduced in Section 3 takes this possibility into account assuming two regimes for GDD. The two periods considered in the steady-state exercise are in line with the two regimes in the nonlinear state space model.

liquidity constraints as they stay unemployed longer. This theory is called the *positive* GDD.

Second, the exit probabilities were, on average, lower after the Great Recession began (blue dashed line) than they were during the pre-recession period (red line). This evidence suggests that either the average exit probability falls or the share of those who have low job-finding rates among newly unemployed workers rises during the economic downturn.

Finally, substantial differences exist in the shape of unemployment-exit probabilities over the duration of unemployment between groups. Specifically, between December 2007 and December 2013, the unemployment hazards of women aged 16 to 24 with a high-school diploma or less increase after one year in unemployment (blue dashed line in the right panel), while those of men aged 25 to 54 with some college or an associate degree do not (blue dashed line in the left panel). This observation indicates the importance of considering different distributions of unobserved heterogeneity and different patterns of GDD for each observable-characteristic group in order to accurately understand the factors that affect the duration distribution of a particular group.

3 Model and Estimation

This section portrays the key empirical analysis. Section 3.1 discusses the identification of the inflows and unemployment-continuation probabilities of workers who have the same observable characteristics but have different unobserved types based on the model of dynamic accounting identity. Section 3.2 introduces the nonlinear state space model that estimates the dynamic latent variables.

3.1 Identification: Dynamic accounting identity for group j

Suppose that there are unemployed individuals with two unobserved types—H and L—in group j.[12] In month t, type H workers have a *high* probability of exiting unemployment next month, denoted by p_{jt}^{H}, and type L workers have a *low* probability of exiting unemployment next month, denoted by p_{jt}^{L}. Each month, there are new inflows of type H and L workers in group j, denoted by w_{jt}^{H} and w_{jt}^{L}, respectively. In addition, type H and L workers' probabilities of exiting unemployment change over the duration of unemployment reflecting GDD. We assume that GDD has limited time-variations, and that the two types of workers are exposed to the same GDD but the GDD may differ by j.

Given this set of assumptions, we can identify the inflows, the unemployment-continuation probabilities of workers with two unobserved types, and the parameters governing GDD with $U_{jt}^{1}, U_{jt}^{2.3}, U_{jt}^{4.6}$, $U_{jt}^{7.12}$, and $U_{jt}^{13.+}$. To provide the intuition of identification, suppose that the economy is in a steady state. I suppress the time subscript, t, accordingly. Assume first that there is no GDD. The sum of w_j^L and w_j^H is U_j^1. The number of individuals unemployed for two months, U_j^2, equals the number of newly unemployed type H and L individuals who continue to stay unemployed for one more month. Likewise, U_j^3 equals the number of newly unemployed type H and L workers who continue to stay unemployed for two more months. The sum of U_j^2 and U_j^3 is $U_j^{2.3}$. By the same token, we can express $U_j^{4.6}$ and $U_j^{7.12}$ as the functions of the two inflows and two unemployment-continuation probabilities. This suggests that we can solve for the four unknowns, w_j^L, w_j^H, p_j^L, and p_j^H, if we observe the four data points, $U_j^1, U_j^{2.3}, U_j^{4.6}$ and $U_j^{7.12}$.

[12] Because we estimate the inflows and unemployment-continuation probabilities of each type, only two types can be identified from the five data points in month t. Considering just the two unobserved types in a given group is a restriction, and incorporating more types is desirable. Nonetheless, assuming the two types of workers is not critical for a few reasons. First, the goal of this analysis is to show that a model with unobserved heterogeneity can go a long way toward explaining the distribution of unemployment duration. Achieving the goal with a simple model suggests that we will reach the same conclusion with a model of more types. Second, in a separate paper by Ahn and Hamilton (Forthcoming), three types are considered to characterize the weekly duration data. However, no statistically significant improvement in the likelihood value is achieved, and the third type tends to converge to one of the two types.

GDD can be jointly identified with the inflows and unemployment-continuation probabilities of two types. Assume for simplicity that GDD is a linear function of τ characterized by one parameter. Then, we can identify the five unknowns—the GDD parameter along with w_j^L, w_j^H, p_j^L, and p_j^H—using the five values $U_j^1, U_j^{2.3}, U_j^{4.6}, U_j^{7.12}$, and $U_j^{13.+}$. Now suppose that we observe $U_j^1, U_j^{2.3}, U_j^{4.6}, U_j^{7.12}$, and $U_j^{13.+}$ in two different periods and GDD does not change over time. We can solve for w_j^L, w_j^H, p_j^L, and p_j^H in each period and use the two remaining data points to characterize GDD. Note that we can characterize nonlinearity in GDD with the two data points. As we consider data from more periods, we can have a more general functional form for GDD. In fact, we use $U_{jt}^1, U_{jt}^{2.3}, U_{jt}^{4.6}, U_{jt}^{7.12}$, and $U_{jt}^{13.+}$ for every month t during the sample period. With such data, we can even allow different regimes for GDD assuming that GDD does not vary over time within a regime.

These steady-state examples illustrate the intuition of identification, but the similar logic applies in a dynamic setting. The four variables, p_{jt}^H, p_{jt}^L, w_{jt}^H, and w_{jt}^L, are dynamic latent variables that change every month. Each group j has a different function for GDD. I use a nonlinear function for GDD to capture a possible non-monotonic pattern of unemployment-exit probabilities over the duration of unemployment. Lastly, following the previous research studying variation in GDD depending on business cycle phases (e.g., Kroft et al. (2013)), I assume that there are two regimes for GDD—low- and high-unemployment-rate regimes—but that the magnitude of GDD does not change within a regime. Given the history of estimated inflows and unemployment-continuation probabilities up to month $t-1$ and the parameters governing GDD, we can estimate the four latent variables in month t that best fit $U_{jt}^1, U_{jt}^{2.3}, U_{jt}^{4.6}, U_{jt}^{7.12}$, and $U_{jt}^{13.+}$. In this way, we can identify the inflows and unemployment-continuation probabilities of workers with unobserved types at each point in time jointly with the GDD parameters based upon the accounting identity for the number of unemployed individuals by the duration of unemployment.[13]

[13]Data on unemployment duration are affected by reporting errors, such as the digit preference documented in Ahn and Hamilton (Forthcoming). In the online appendix, I elaborate on why this issue is not relevant for

3.2 Nonlinear state space model

The model of dynamic accounting identity of unemployment for group j is cast into a nonlinear state space model. The measurement equation is the following. In month t, the number of individuals with observable characteristic j who have been unemployed for one month, U^1_{jt}, is the sum of newly unemployed type H and L workers, w^H_{jt} and w^L_{jt}, respectively.

$$U^1_{jt} = w^H_{jt} + w^L_{jt}$$

I assume that for unemployed individuals in group j who have already been unemployed for τ months as of time $t-1$, the fraction of those who will still be unemployed at t is given by

$$p^z_{jt}(\tau) = \exp[-\exp(x^z_{jt} + d^{g_t}_{j\tau})]$$

for $z = H, L$. The notation x^z_{jt} is a time-varying parameter determining the unemployment-continuation probability for workers of type z in group j regardless of their duration, and it captures cross-sectional heterogeneity in unemployment-continuation probabilities between the two types. Because type H workers have a higher exit probability from unemployment, I set $x^H_{jt} > x^L_{jt}$.

The term $d^{g_t}_{j\tau}$ captures the effect from GDD. To characterize the possible nonlinearity in the relations between the unemployment-exit probability and the unemployment duration as illustrated in Section 2, I use a linear spline for $d^{g_t}_{j\tau}$ with breaks at $\tau = 6$ and 12. One source of the nonlinearity is the exhaustion of unemployment insurance (UI) benefits. As shown by Katz and Meyer (1990), unemployed individuals tend to exit unemployment more rapidly, before they exhaust their UI benefits. This can create positive GDD in the first 6 months in unemployment, considering the maximum duration of UI benefits is 6 months in normal times. After 6 months in unemployment, workers might become discouraged or

the purpose of the present study and which alternative strategies may be used in further research.

discriminated against by potential employers because of their long jobless spells, creating negative GDD. In addition, after having been jobless long enough, unemployed workers might quit job search or take a low-paying job, which creates the positive GDD. To capture this possibility, I consider another break at $\tau = 12$.

In addition, when a state's unemployment rate is higher than 6.5%, the UI benefits are extended up to 52 weeks, suggesting the pattern of GDD can change over time. Therefore, I allow $d_{j\tau}^{g_t}$ to have two regimes depending on the level of unemployment rate at time t. I set $g_t = E$ (*Expansion*) for month t when the unemployment rate is lower than 6.5% and $g_t = R$ (*Recession*) when the unemployment rate is 6.5% or higher.[14] Lastly, I set $d_{j\tau}^{g_t}$ to be constant for $\tau \geq 24$, assuming that those who have been unemployed for 2 years or longer are largely the same in their probabilities of exiting unemployment.

The functional form of $d_{j\tau}^{g_t}$ is as follows:

$$d_{j\tau}^{g_t} = \begin{cases} \delta_{j1}^{g_t}(\tau - 1) & \text{for } \tau < 6 \\ \delta_{j1}^{g_t}[(6-1)-1] + \delta_{j2}^{g_t}[\tau - (6-1))] & \text{for } 6 \leq \tau < 12 \\ \delta_{j1}^{g_t}[(6-1)-1] + \delta_{j2}^{g_t}[(12-1)-(6-1)] + \delta_{j3}^{g_t}[\tau - (12-1)] & \text{for } 12 \leq \tau < 24 \\ \delta_{j1}^{g_t}[(6-1)-1] + \delta_{j2}^{g_t}[(12-1)-(6-1)] + \delta_{j3}^{g_t}[(24-1)-(12-1)] & \text{for } 24 \leq \tau. \end{cases}$$

Positive values of $\delta_{jh}^{g_t}$ for $h = 1, 2, 3$ capture positive GDD, while negative values capture negative GDD.

Let $P_{jt}^{z}(k)$ be the fraction in group j of individuals of type z who were unemployed for one month as of date $t-k$ and are still unemployed at t. Then, $P_{jt}^{z}(k)$ is written as a product of monthly fractions $p_{j,t-k+\tau}^{z}(\tau)$ for $\tau = 1, 2, ..., k$ as follows:

$$P_{jt}^{z}(k) = p_{j,t-k+1}^{z}(1)p_{j,t-k+2}^{z}(2)...p_{jt}^{z}(k).$$

[14]For robustness checks, I additionally consider the third regime with the threshold unemployment rate at 8.0%. Nonetheless, there are no material changes in the estimates, and the likelihood values are not substantially improved.

Note that individuals unemployed for two to three months at t include those who become newly unemployed at time $t-1$ and look for a job at t, and those who become newly unemployed at $t-2$ and continue to look for a job at $t-1$ and t. Thus, the number of individuals in group j unemployed for two to three months in month t, $U_{jt}^{2.3}$, is written as follows:

$$U_{jt}^{2.3} = \sum_{z=H,L} \left[w_{j,t-1}^{z} P_{jt}^{z}(1) + w_{j,t-2}^{z} P_{jt}^{z}(2) \right].$$

Likewise, the number of those who have been unemployed for between m_1 and m_2 months at time t ($U_{jt}^{m_1.m_2}$) is

$$U_{jt}^{m_1.m_2} = \sum_{z=H,L} \sum_{k=m_1-1}^{m_2-1} \left[w_{j,t-k}^{z} P_{jt}^{z}(k) \right].$$

I further assume that each data point, $U_{jt}^{1}, U_{jt}^{2.3}, U_{jt}^{4.6}, U_{jt}^{7.12}$ and $U_{jt}^{13.+}$, is observed with a measurement error, $r_{jt}^{1}, r_{jt}^{2.3}, r_{jt}^{4.6}, r_{jt}^{7.12}$ and $r_{jt}^{13.+}$, respectively:

$$\begin{aligned}
U_{jt}^{1} &= w_{jt}^{H} + w_{jt}^{L} + r_{jt}^{1} \\
U_{jt}^{2.3} &= \sum_{z=H,L} \left[w_{j,t-1}^{z} P_{jt}^{z}(1) + w_{j,t-2}^{z} P_{jt}^{z}(2) \right] + r_{jt}^{2.3} \\
U_{jt}^{4.6} &= \sum_{z=H,L} \sum_{k=3}^{5} \left[w_{j,t-k}^{z} P_{jt}^{s}(k) \right] + r_{jt}^{4.6} \\
U_{jt}^{7.12} &= \sum_{z=H,L} \sum_{k=6}^{11} \left[w_{j,t-k}^{z} P_{jt}^{z}(k) \right] + r_{jt}^{7.12} \\
U_{jt}^{13.+} &= \sum_{z=H,L} \sum_{k=12}^{47} \left[w_{j,t-k}^{z} P_{jt}^{z}(k) \right] + r_{jt}^{13.+}.
\end{aligned}$$

I terminate the calculations after 4 years of unemployment. We can define the likelihood function for the observed data conditional on state variables by assuming that the vector of measurement errors $r_{jt} = [r_{jt}^{1}, r_{jt}^{2.3}, r_{jt}^{4.6}, r_{jt}^{7.12}, r_{jt}^{13.+}]'$ is independent Normal,

$$r_{jt} \sim N(0, R_j)$$

with $R_j = diag((R_j^1)^2, (R_j^{2.3})^2, (R_j^{4.6})^2, (R_j^{7.12})^2, (R_j^{13.+})^2)$ where R_j^1, $R_j^{2.3}$, $R_j^{4.6}$, $R_j^{7.12}$, and $R_j^{13.+}$ are the standard deviations of $r_{jt}^1, r_{jt}^{2.3}, r_{jt}^{4.6}, r_{jt}^{7.12}$, and $r_{jt}^{13.+}$, respectively.

Let me turn to the state equation. Let ξ_{jt} be the vector $[w_{jt}^L, w_{jt}^H, x_{jt}^L, x_{jt}^H]'$, and ϵ_{jt} be the vector $[\epsilon_{jt}^{Lw}, \epsilon_{jt}^{Hw}, \epsilon_{jt}^{Lx}, \epsilon_{jt}^{Hx}]'$. The assumption that the latent factors evolve as random walks would be written as

$$\xi_{jt} = \xi_{j,t-1} + \epsilon_{jt}, \qquad \epsilon_{jt} \sim N(0, \Sigma_j)$$

with $\Sigma_j = diag((\sigma_{jL}^w)^2, (\sigma_{jH}^w)^2, (\sigma_{jL}^x)^2, (\sigma_{jH}^x)^2)$ where $\sigma_{jL}^w, \sigma_{jH}^w, \sigma_{jL}^x$, and σ_{jH}^x are the standard deviations of ϵ_{jt}^{Lw}, ϵ_{jt}^{Hw}, ϵ_{jt}^{Lx}, and ϵ_{jt}^{Hx}, respectively.

Because the measurement equation is a function of $\{\xi_{jt}, \xi_{j,t-1}, ..., \xi_{j,t-47}\}$, the joint distribution of ξ_{jt} from $t-47$ to t is captured by the state equation as follows:

$$\underbrace{\begin{bmatrix} \xi_{jt} \\ \xi_{j,t-1} \\ \xi_{j,t-2} \\ \vdots \\ \xi_{j,t-47} \end{bmatrix}}_{192\times 1} = \underbrace{\begin{bmatrix} \underbrace{I}_{4\times 4} & \underbrace{0}_{4\times 4} & 0 & 0 & ... & 0 & 0 & 0 \\ I & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & I & 0 & 0 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & ... & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & 0 & I & 0 \end{bmatrix}}_{192\times 192} \underbrace{\begin{bmatrix} \xi_{j,t-1} \\ \xi_{j,t-2} \\ \xi_{j,t-3} \\ \vdots \\ \xi_{j,t-48} \end{bmatrix}}_{192\times 1} + \underbrace{\begin{bmatrix} \underbrace{\epsilon_{jt}}_{4\times 1} \\ \underbrace{0}_{4\times 1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{192\times 1}$$

where I and 0 denote a (4×4) identity and zero matrix.

The model is a nonlinear state space model where the measurement equation is nonlinear in the latent variables of interest. Therefore, the extended Kalman filter is used to form the likelihood function for the observed data and make an inference on the dynamic latent variables. The model has 15 parameters to estimate for each group j—namely, the diagonal terms in the variance matrices Σ_j and R_j, and the parameters governing GDD, δ_{j1}^E, δ_{j2}^E , δ_{j3}^E, δ_{j1}^R, δ_{j2}^R and δ_{j3}^R. The system of equations is estimated with maximum likelihood. I report full-sample smoothed inferences, denoted by $\hat{\xi}_{jt|T}$.[15]

4 Empirical results

This section reports the estimation results. I discuss the unemployment-continuation probabilities of types H and L in Section 4.1 and the two inflows in Section 4.2. Among the various worker characteristics I consider—gender, age, education, industry, occupation, and reason for unemployment—I particularly focus on "reason for unemployment" for two reasons. First, the estimates indicate that the single observed worker characteristic that is most similar to the type L attribute is permanent separation—one of the reasons for unemployment. Second, previous studies such as Fujita and Moscarini (2017) and Hall and Schulhofer-Wohl (2018) also argue that the reason for unemployment is the key worker characteristic determining differences in the job-search outcomes of unemployed individuals.[16]

4.1 The unemployment-continuation probabilities of two unobserved types

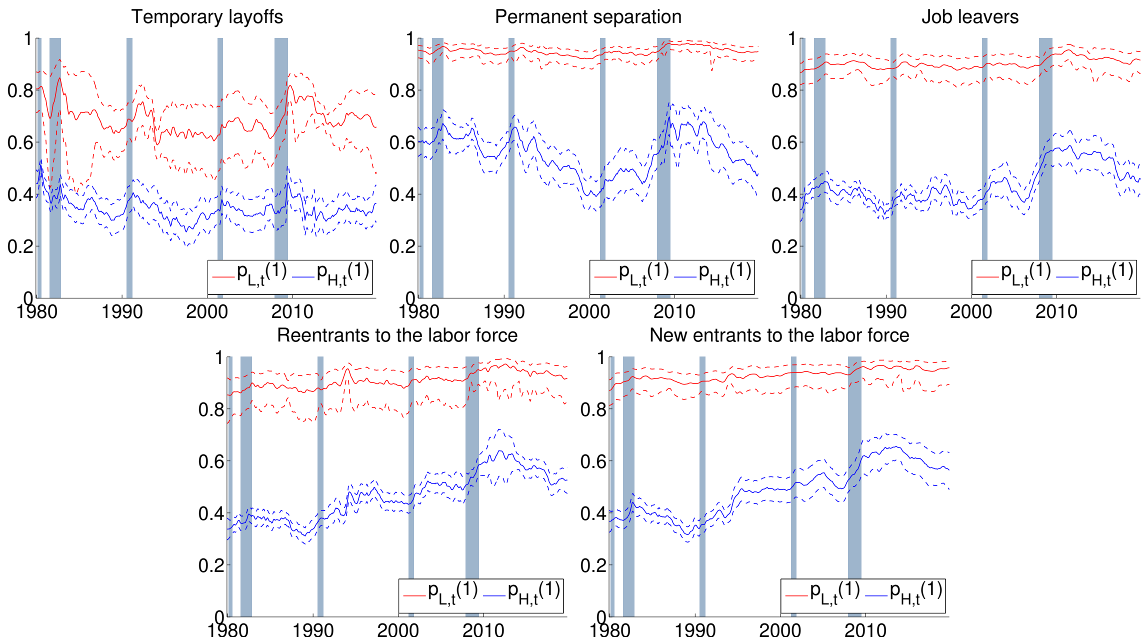
In this section, I first show the unemployment-continuation probabilities of type H and L workers without GDD to examine the magnitude of difference in unemployment hazards between two unobserved types at each point in time. Next, I report GDD parameters and analyze the role of GDD in the duration dependence of unemployment hazards. Last, I explore what these estimates tell us about changes in long-term unemployment.

The unemployment-continuation probabilities of newly unemployed type H and L individuals are reported in Figure 3. Note that these probabilities are not affected by GDD. It is commonly observed across the five groups that the difference in probabilities between the two types is substantial. Average type L continuation probabilities are between 0.69 and

[15] More details about the estimation algorithm are found in the online appendix.

[16] The estimation results for the groups of other observable characteristics are documented in the online appendix. Consistently, the variance of unemployment attributed by the inflows and unemployment-continuation probabilities exhibits the largest difference across groups, when the data are disaggregated by reason for unemployment.

0.95, while average type H continuation probabilities are between 0.34 and 0.56 as summarized in Table 1. The gap between type H and L probabilities ranges between 0.35 and 0.46.



Note: Shaded areas denote NBER recessions. The dashed lines denote 90% confidence intervals.
Source: Author's calculation.

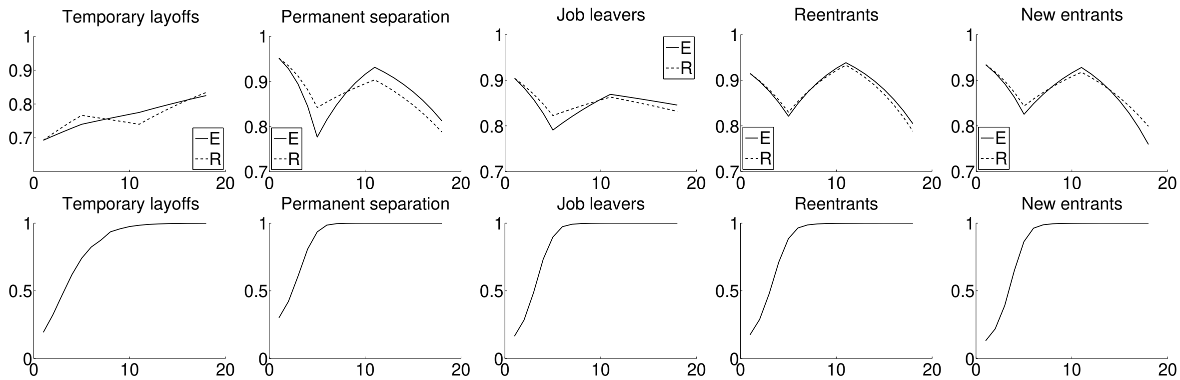
Figure 3: Probability that a newly unemployed worker of each type will still be unemployed the following month by reason for unemployment.

	TL	PS	JL	RE	NE
Type L continuation probability	0.69	0.95	0.90	0.91	0.93
Type H continuation probability	0.34	0.56	0.44	0.47	0.49
Type L share in inflows	0.19	0.31	0.15	0.17	0.12

Note: TL, PS, JL, RE, and NE stand for temporary layoffs, permanent separations, job leavers, reentrants to the labor force, and new entrants to the labor force, respectively.

Table 1: Average inflows and unemployment-continuation probabilities by reason for unemployment (1980-2019)

The unemployment-continuation probabilities of both types H and L tend to go up during an economic downturn. During the Great Recession and its aftermath, the two probabilities reach the highest or close to the highest levels since 1980 in all groups.



Note: In the upper panels, the variation in type L unemployment-continuation probability reflects effects from GDD. The solid line denotes regime "E" and the dashed line denotes regime "R."
Source: Author's calculation.

Figure 4: Average type L unemployment-continuation probability (upper panels) and average type L share (lower panels) over the duration of unemployment

Overall, GDD parameters are positive for the first 6 months of unemployment, negative between 7 and 12 months, and positive again after 1 year as shown in the upper panels of Figure 4.[17] GDD parameters are largely statistically significant.[18] The lower panels of Figure 4 show that type L share increases rapidly as the unemployment duration progresses up to 6 months. The positive GDD and the rapid rise in type L share in the range of duration with 6 months or less indicate that unobserved heterogeneity is the main explanation for the negative duration dependence, considering that a large part of the deterioration in unemployment hazards takes place in the first 6 months of unemployment.[19][20]

[17] I plot the unemployment-continuation probability of type L workers over unemployment duration because the share of type H workers becomes close to zero after 6 months in unemployment, as shown by the lower panels of Figure 4. The positive GDD in long-term unemployment might reflect that unemployed workers become more likely to quit looking for a job due, for instance, to discouragement from an unsuccessful job search. Meanwhile the negative GDD might reflect firms' discrimination against the long-term unemployed. It is possible that negative GDD is prevalent in transitions from unemployment to employment, while positive GDD is in effect in transitions from unemployment to nonparticipation.

[18] The parameter estimates are documented in the online appendix.

[19] Kroft et al. (2013) show that the negative duration dependence is observed in the first eight months of the job search, and Jarosch and Pilossoph (2018) argue that this pattern can be explained by unobserved worker heterogeneity.

[20] The rising type L share over the duration of unemployment captures the dynamic sorting that arises from

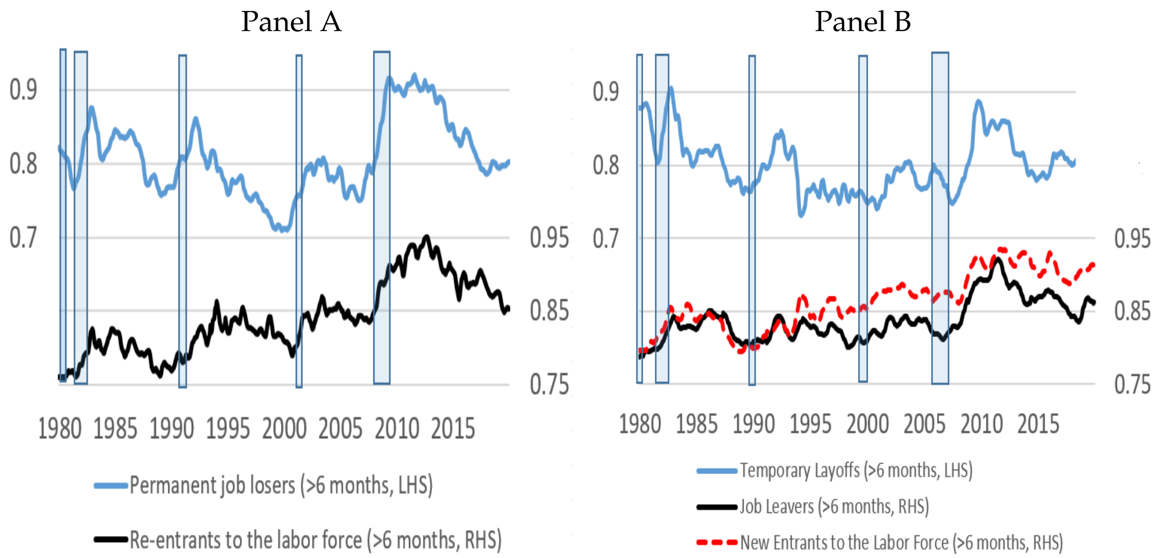
In addition, GDD has a small effect on the mean duration of unemployment. In short-term unemployment with duration 6 months or less, the positive GDD lowers the mean duration by 0.3 month, on average. In long-term unemployment with duration longer than 6 months, the negative GDD pushes up the mean duration by 2.5 months and the positive GDD lowers the average duration by 2.2 months, raising the mean duration of unemployment by 0.3 months on net.[21] Changes in the regime of GDD do not significantly alter the contribution of GDD to the mean duration, either.[22]

What do the results imply about long-term unemployment? The estimates suggest that type L workers' unemployment-continuation probabilities are an important determinant of changes in long-term unemployment, as type L workers mainly constitute long-term unemployment. In fact, the unemployment-continuation probabilities of long-term type L workers stay at high levels for a few years after the Great Recession as shown in Figure 5, coinciding with the period when the mean duration of unemployment continues to rise to a record-high level. In addition, the limited recoveries in the mean duration of unemployment after the 2001 and 2007 recessions are also closely associated with type L permanent

the difference in unemployment hazards between the two types. Type L workers take a larger share in long-term unemployment, as they are less likely to leave unemployment status. Figure 4 shows that the dynamic sorting is the key contributor to the negative duration dependence in unemployment hazards. However, this finding does not necessarily mean that the compositional shift of type L workers should be a dominant factor in the rise of the mean duration of unemployment. Note that two circumstances can raise the mean duration of unemployment: when type L workers stay unemployed longer than before and when the type L share in the unemployment pool rises because more type L workers lose jobs than than type H workers do. When type L workers' unemployment duration becomes longer during a recession, the type L duration can be the key factor in the rise of the mean duration and the dynamic sorting can still be the key contributor to the negative duration dependence. Therefore, the potentially greater importance of type L duration in accounting for the rise in the mean duration of unemployment does not contradict the importance of dynamic sorting in the negative duration dependence of unemployment hazards.

[21] The contribution of positive GDD in the short-term unemployment is computed from the difference between the mean duration where $d^{g_t}_{j\tau} = d^{g_t}_{j5}$ for $\tau < 6$ and that with the full path of GDD (the baseline). The contribution of negative GDD is computed from the difference between the mean duration where only the negative GDD is in action in the long-term unemployment and the average duration where GDD does not change from $\tau \geq 6$. The former is calculated by fixing $d^{g_t}_{j\tau}$ for $\tau \geq 12$ at $d^{g_t}_{j,11}$, and the latter is computed by fixing $d^{g_t}_{j\tau}$ for $\tau \geq 6$ at $d^{g_t}_{j5}$. The difference between the baseline and the former counterfactual is the contribution from positive GDD in the long-term unemployment.

[22] On average, GDD raises the mean duration by 0.3 month on net in the E regime, and lowers it by 0.2 month on net in the R regime.



Note: Refer to the left axis for the blue line and the right axis for the black and red lines.
Source: Author's calculation.

Figure 5: Unemployment-continuation probability of type L individuals unemployed for longer than 6 months by reason for unemployment

job losers and reentrants to the labor force. Their long-term unemployment-continuation probabilities did not return to the pre-recession levels after the two recessions as shown in Panel A of Figure 5.[23]

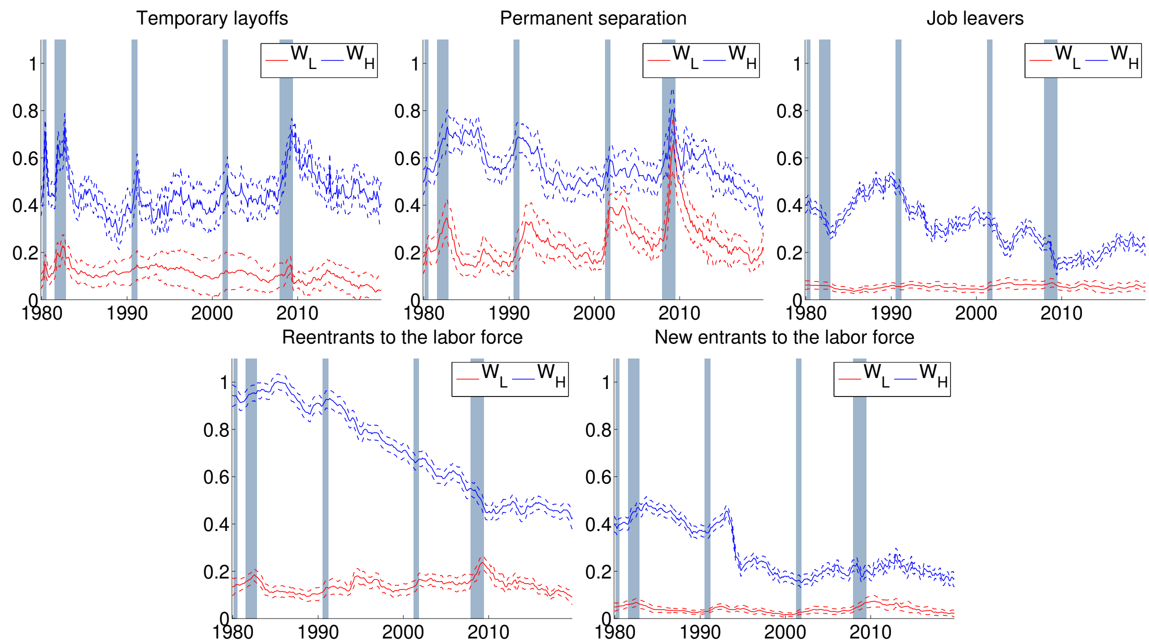
4.2 The inflows of two unobserved types

In this section, I first report the estimated inflows of type H and L workers. Next, I explore which group constitutes the majority of total type L inflows and mainly drives the countercyclicality of aggregate type L inflows to identify the observable worker characteristic that is most closely associated with the type L attribute. Lastly, I briefly discuss the trends in the inflows.

The smoothed estimates of type H and L newly unemployed individuals are displayed in Figure 6. Type L individuals make up a small portion of the inflows and represent, on

[23]Section 6 explores the drivers of the trends in probability estimates in more detail.

average, 13% to 28% of newly unemployed individuals across all groups (Table 1). Notably, permanent job losers have the largest type L share among the newly unemployed.[24]



Note: Units are in hundred thousand individuals. Shaded areas denote NBER recessions. The dashed lines denote 90% confidence intervals.
Source: Author's calculation.

Figure 6: Number of newly unemployed workers of each type by reason for unemployment.

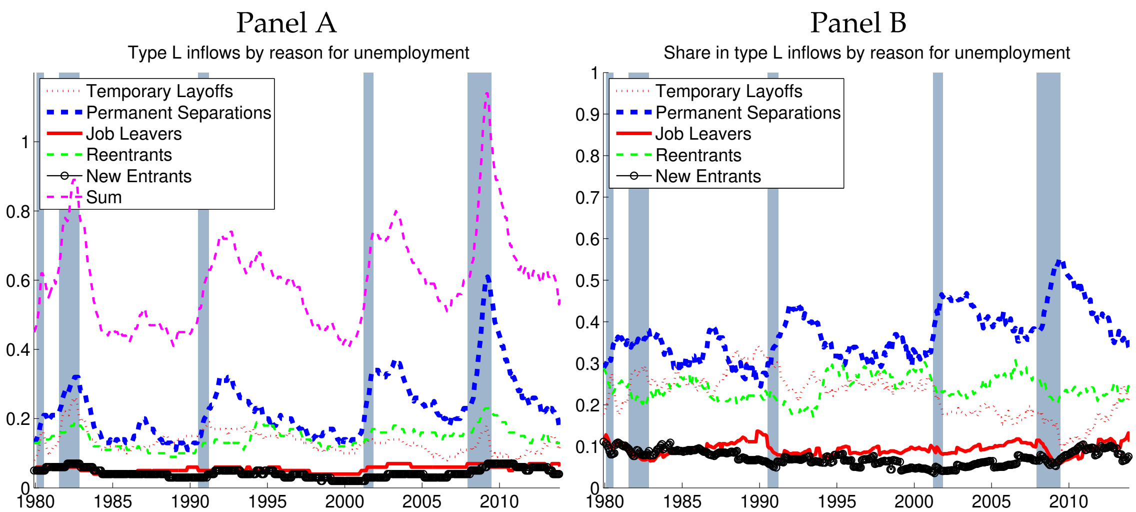
The dynamic features of the inflows are quite different both between the unobserved types and across the five groups. The inflows of job losers are countercyclical. Notably, type L inflows of permanent job losers exhibit the strongest countercyclicality among all the inflows and essentially drive the countercyclical variation in total type L inflows.[25] However, the inflows of those who did not indicate job loss as their reason for unemployment show mixed cyclicality. The inflows of type L reentrants to the labor force exhibit weak counter-

[24] There are unemployed individuals who indicate temporary layoffs as their reason for unemployment but change their answers to permanent separation in the subsequent months. The model does not take these transitions into account. In the online appendix, I explain in detail why this problem does not affect the results of my analysis.

[25] This point is summarized in Panel A of Figure 7. No groups with other observable characteristics drive the rise of type L inflows during a recession as much as permanent job losers do. Figures analogous to Figure 7 for other worker characteristics are documented in the online appendix.

cyclicality, those of type H job leavers are procyclical, and the rest are largely acyclical.

Due mainly to type L permanent job losers, the countercyclicality of total type L inflows is greater than that of total type H inflows. This observation indicates that disproportionately more workers with low reemployment prospects flow into the unemployment pool during an economic downturn.



Note: TL, PS, JL, RE, and NE stand for temporary layoffs, permanent separations, job leavers, reentrants to the labor force, and new entrants to the labor force, respectively. For Panel A, units are in million individuals. Source: Author's calculation.

Figure 7: Composition of type L inflows by reason for unemployment.

More importantly, all these observations indicate that permanent separation is likely the observable characteristic that is most closely associated with the type L attribute.[26] As shown in the composition of type L inflows by reason for unemployment (Panel B of Figure 7), permanent job losers take the largest share, 45%, among type L inflows. At the same time, the share of type L workers in the inflows is the largest among permanent job losers (Table 1).

[26] Note that the results should be interpreted with caution. I do not claim that type L workers are permanent job losers or permanent job losers are type L. Among worker characteristics observable in the CPS, permanent job loss is the observable worker characteristic that is most closely correlated with the type L attribute. As shown in Figure 3, there still exist a significant difference in the unemployment-exit probabilities among permanent job losers, which is the main theme of the paper.

Lastly, it is notable that type H inflows of job leavers and reentrants to the labor force trend down throughout the sample period, while their type L inflows and the inflows of other groups do not exhibit any particular trend.[27] This observation suggests the composition of inflows gradually shifts toward type L, which may have driven up the mean duration of unemployment over time.[28]

5 Contribution of worker heterogeneity to the evolution of unemployment duration

Using the model's estimates, this section analyzes how much the compositional shift of workers with observable and unobservable attributes, as well as changes in the unemployment duration of each type in each group, accounts for the evolution of the aggregate duration of unemployment. I focus on the period from December 2007 to December 2011, and the entire sample period from January 1980 to December 2019 to examine the role of each factor in the cyclical and low-frequency variations in the mean duration of unemployment.

Let D_{jt}^{z} denote the mean duration of unemployment of workers in group j with type z for $z = H, L$. The object, D_{jt}^{z}, is computed from the following:

$$D_{jt}^{z} = \frac{\sum_{k=1}^{48} \left[w_{j,t-k+1}^{z} P_{jt}^{z}(k-1) \right] k}{\sum_{k=1}^{48} \left[w_{j,t-k+1}^{z} P_{jt}^{z}(k-1) \right]},$$

where k denotes the number of months in unemployment and $P_{jt}^{z}(0) = 1$.

[27] Type H inflows of new entrants to the labor force step down in 1994 due mainly to the CPS redesign in 1994 that broadens the range of individuals who are classified as reentrants to the labor force and narrows the range of new entrants to the labor force. Polivka and Miller (1998) provide adjustment factors for the unemployment rates of the two groups that are methodologically consistent but do not provide factors for their distribution of unemployment duration. Therefore, I do not adjust the duration distribution of new entrants and reentrants to the labor force after 1994. Even if I increase the type H inflows of new entrants with the adjustment factors for the unemployment rates, the inflows do not exhibit any particular trend.

[28] More detailed analyses on the trends in inflows are provided in Section 6.

Let f_{jt}^{z} denote the fraction of type z workers in the unemployment of group j. The object, f_{jt}^{z}, is calculated from

$$f_{jt}^{z} = \frac{\sum_{k=1}^{48}\left[w_{j,t-k+1}^{z} P_{jt}^{z}(k-1)\right]}{\sum_{z=H,L}\sum_{k=1}^{48}\left[w_{j,t-k+1}^{z} P_{jt}^{z}(k-1)\right]}.$$

The mean duration of unemployment, denoted as D_t^{full}, is written into the following:

$$D_t^{full} = \sum_{j=1}^{J} F_{jt}(f_{jt}^{H} D_{jt}^{H} + f_{jt}^{L} D_{jt}^{L}),$$

where F_{jt} is the fraction of group j in the aggregate unemployment.

The contribution from a compositional shift of workers with observable characteristics is calculated by letting only F_{jt} vary over time, while fixing D_{jt}^{z} and f_{jt}^{z} at the values observed at the beginning of the period of analysis, t_0. Let $D_{jt_0}^{z}$ and $f_{jt_0}^{z}$ be the unemployment duration and the fraction of type z workers in group j in month t_0, respectively. Then the mean duration of unemployment explained by shifts in the composition of groups with observable characteristics, denoted as D_t^{o}, is

$$D_t^{o} = \sum_{j=1}^{J} F_{jt}(f_{jt_0}^{H} D_{jt_0}^{H} + f_{jt_0}^{L} D_{jt_0}^{L}).$$

The contribution from the compositional shift of workers with unobserved types is calculated by letting f_{jt}^{H} and f_{jt}^{L} vary over time while fixing the other components at the values of t_0. Then the mean duration of unemployment accounted for by the compositional shift of workers with unobserved types, denoted as D_t^{u}, is

$$D_t^{u} = \sum_{j=1}^{J} F_{jt_0}(f_{jt}^{H} D_{jt_0}^{H} + f_{jt}^{L} D_{jt_0}^{L}).$$

Next, consider the case in which only the duration of type L workers in group j is allowed to vary over time. Then the predicted mean duration of unemployment, denoted as D_{Lt}, is written into

$$D_{Lt} = \sum_{j=1}^{J} F_{jt_0}(f^H_{jt_0} D^H_{jt_0} + f^L_{jt_0} D^L_{jt}).$$

Likewise, by allowing only D^H_{jt} to change, the predicted mean duration, denoted as D_{Ht}, is expressed as follows:

$$D_{Ht} = \sum_{j=1}^{J} F_{jt_0}(f^H_{jt_0} D^H_{jt} + f^L_{jt_0} D^L_{jt_0}).$$

Changes in the regime of GDD affect D_{Ht} and D_{Lt} through D^H_{jt} and D^L_{jt}, respectively.[29] Let D^E_{Ht} (D^E_{Lt}) denote the predicted mean duration at t driven only by changes in type H (L) duration, when the GDD parameters are in regime E. The effect of changes in the regime on the mean duration of unemployment, denoted as G_t, is calculated from the following:

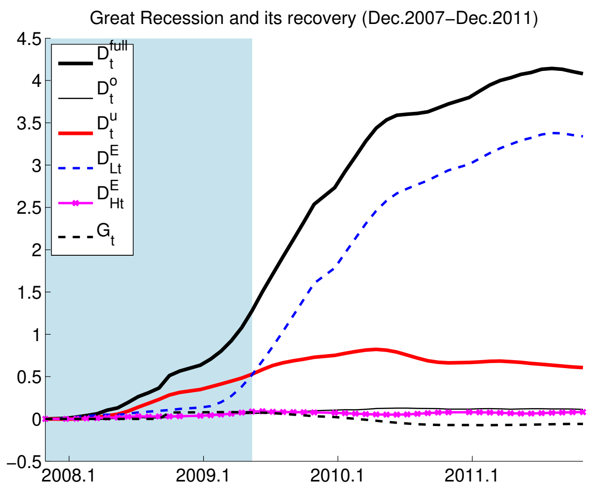
$$G_t = (D_{Ht} - D^E_{Ht}) + (D_{Lt} - D^E_{Lt}).$$

Figure 8 plots the paths of D^o_t(black line), D^u_t(red line), D^E_{Lt}(blue dashed line), D^E_{Ht}(pink line with x's), and G_t (black dashed line) between December 2007 and December 2011.[30] Note that the values in December 2007 are set to zero.[31] The unemployment duration of type L workers (blue dashed line) is the most important contributor to the rise in the mean duration of unemployment, accounting for about 80% of the rise. This factor begins to raise the mean duration in the latter part of the Great Recession, and continues to drive it upward to the unprecedentedly high level in 2011. The remaining increase in mean duration is explained by the compositional shift of unobserved types. The increased share of type L

[29] Note that the GDD parameters are time-invariant within a regime. Therefore, changes in the distribution of unemployment duration are driven by type H and L inflows, their unemployment-continuation probabilities, or changes in GDD.

[30] In 2011, the predicted mean duration, D^{full}_t, is about one month shorter than in the data. The model attributes the difference to measurement errors in the duration data. Farber et al. (2015) and Ahn and Hamilton (Forthcoming) show evidence that measurement errors in long-term unemployment likely become substantially larger during this period. The model fits the data very well before the Great Recession.

[31] More specifically, D^{full}_{t0} is subtracted from D^o_t, D^u_t, D^E_{Lt}, D^E_{Ht}, and G_t.



Source: Author's calculation.

Figure 8: Contribution of each factor to changes in the mean duration of unemployment from the level in December 2007 (units are in months).

workers (solid red line) is the key driver of the rise in unemployment duration during the Great Recession, but its role becomes limited in the post-recession period when the mean duration continued to rise. Meanwhile, the shift in the share of workers along observable characteristics (thin black line), the unemployment duration of type H workers (pink line with circle) and changes in the regime of GDD (black dashed line) provide little contribution to the evolution of unemployment duration during this period.[32]

I also analyze the contribution of each factor to changes in the mean duration of unemployment between January 1980 and December 2019.[33] As shown in Panel A of Figure 9, the average duration of unemployment doubles from 10.4 weeks in January 1980 to 20.8 weeks in December 2019. Panel B displays how much D_t^o, D_t^u, D_{Lt}, and D_{Ht} account for the rise during the four decades with the value in January 1980 set to zero. Over the whole period, the increased share of type L workers (solid red line) accounts for 30% of the rise, and the

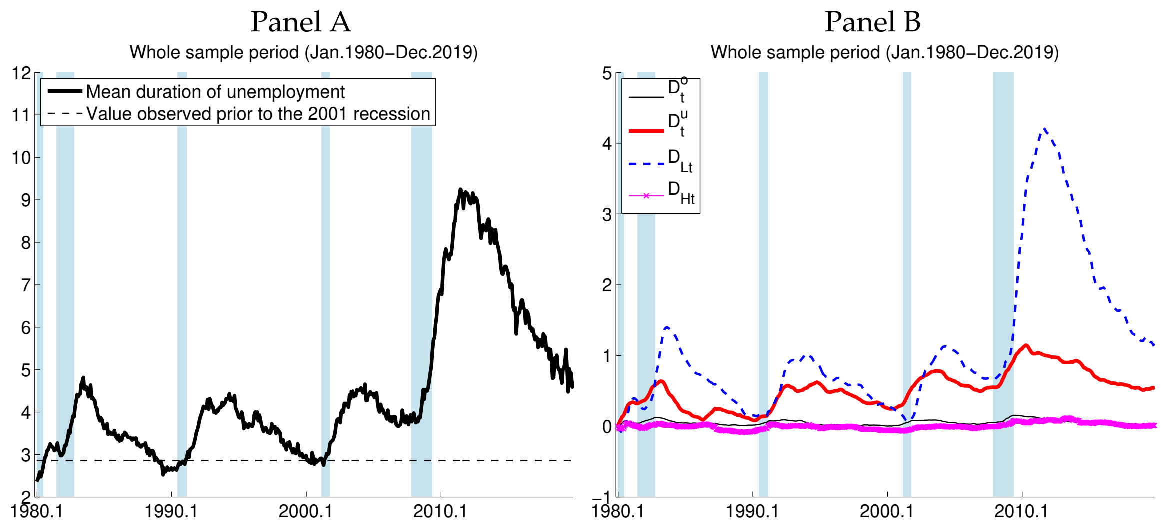
[32]The result indicating the little role in the compositional shift of workers with observable characteristics is consistent with the findings of Krueger et al. (2014) and Kroft et al. (2016).

[33]I do not separately report the contribution from changes in the GDD, because the estimated GDD parameters do not have trends. To my knowledge, no previous studies investigated low-frequency changes in GDD.

lengthier duration of type L workers (dashed blue line) explains 65%. Both the compositional shift of workers with unobserved types and the unemployment duration of type L workers play crucial roles in the secular rise in the mean duration of unemployment.

More specifically, the type L share (solid red line) begins to drive up the mean duration of unemployment from the 1990s. This is mainly attributed to the downward trends in the type H inflows of job leavers and reentrants to the labor force. Quite differently, the contribution from type L duration to the rising trend (blue dashed line) emerges from the 2000s. After the 2001 recession, type L duration begins to impede the full recovery of the mean duration to pre-recession levels. By the end of 2019, 10 years into the recovery after the Great Recession, the mean duration of unemployment predicted solely by type L duration is still higher than the level before the recession. This is mainly due to the limited recovery of type L unemployment-continuation probabilities of permanent job losers and reentrants to the labor force who constitute about 70% of type L unemployment. Meanwhile, the contributions from type H duration (pink line with circle) and the compositional shift of workers with observable worker characteristics (thin black line), again, are not important.

All told, the duration of type L workers is the most important contributor to both cyclical and low-frequency variations in the mean duration of unemployment. The compositional shift of workers with unobserved types is the secondary factor but still explains the bulk of variation. Meanwhile, the contributions from observed heterogeneity and type H duration are negligible. These results suggest unobserved heterogeneity is crucial in the dynamics of the aggregate duration of unemployment.

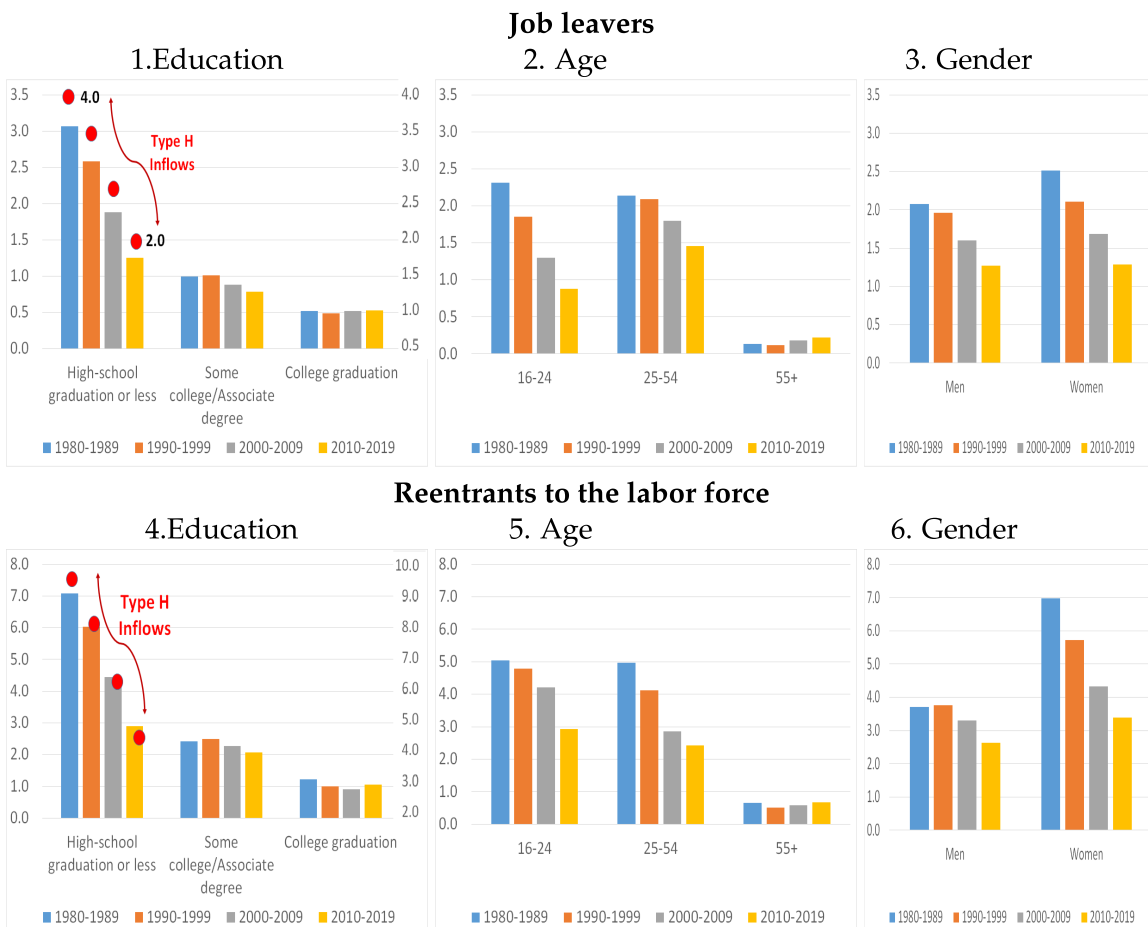


Source: Author's calculation.

Figure 9: Contribution of each factor to changes in the mean duration of unemployment from the level in January 1980 (units are in months).

6 What explains the trends in type H inflows and type L unemployment-continuation probabilities?

The empirical results in the previous section show that the secular rise in the mean duration of unemployment is mainly driven by two factors. First, type H inflows of job leavers and reentrants to the labor force trend down, driving the secular decrease in total type H inflows. Second, type L unemployment-continuation probabilities of permanent job losers and reentrants do not fully recover to the pre-recession levels from the 2000s.[34] I explore the driver of the former in Section 6.1 and that of the latter in Section 6.2.



Note: Units are in hundred thousand individuals. The right axes in Panels 1 and 4 are for type H inflows (red dots).
Source: Author's calculation.

Figure 10: Number of newly unemployed individuals by education, age, and gender among job leavers and reentrants to the labor force.

6.1 The source of downtrend in type H inflows of job leavers and reentrants

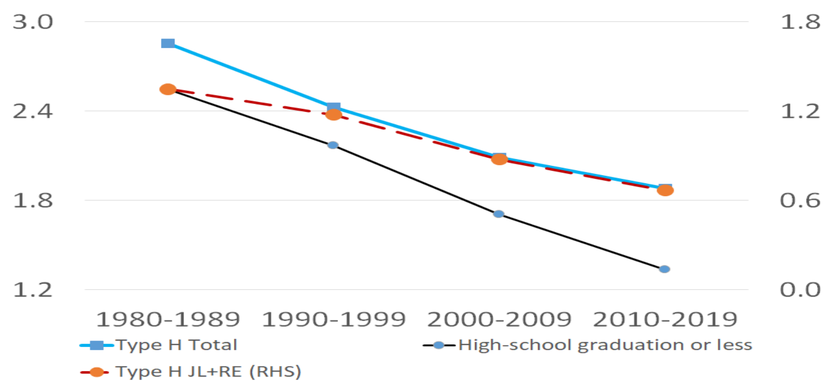
What drives the downtrend in type H inflows of job leavers and reentrants to the labor force? To answer this question, I decompose newly unemployed job leavers and reentrants to the labor force by gender, age, and educational attainment (Figure 10). Consider job leavers first. Between the 1980s and 2010s, the monthly inflows of type H job leavers drop by 200 thousands per month (Panel 1), while those of type L workers are largely unchanged.[35] Similar to the decline in type H job leavers' inflows, the monthly inflows of job leavers whose educational attainment is less than or equal to high-school graduation drop by 180 thousands, from 300 thousands per month in the 1980s to 120 thousands per month in the 2010s. Meanwhile, those with higher education stay flat, similar to the inflows of type L job leavers. However, it is hard to pin down the group whose inflows drop as much as type H inflows do, when decomposed by age and gender (Panels 2 and 3). All told, among job leavers, the secular decrease in type H inflows is mainly driven by less-educated workers.

The association between the type H attribute and lower education is also observed among reentrants to the labor force (lower panels in Figure 10). Between the 1980s and 2010s, the decline of type H inflows is close to that of individuals whose educational attainment is high-school graduation or less (Panel 4). Meanwhile, both their type L inflows and the inflows of reentrants whose education level is higher than high-school graduation are largely flat. Again, the similar dichotomy is not observed when decomposing the data by age and gender (Panels 5 and 6).[36] To summarize, among reentrants to the labor force, the downtrend in type H inflows is also accounted for by the decreased inflows of less educated

[34] The unemployment-continuation probabilities of type H reentrants and new entrants to the labor force also show a rising trend, which likely reflects the increased labor force attachment of women. As it contributes little to the secular rise in the mean duration of unemployment, I do not discuss the uptrend and its source in the main text. Related discussion is found in the online appendix.

[35] Type L inflows of job leavers are 50 thousands (1980-1989), 54 thousands (1990-1999), 62 thousands (2000-2009), and 54 thousands (2010-2019).

[36] Type H inflows of reentrants to the labor force are 0.94 million (1980-1989), 0.82 million (1990-1999), 0.61 million (2000-2009), and 0.46 million (2010-2019). Their type L inflows during the same periods are 0.13, 0.14, 0.16, and 0.14 million.



Note: Units are in million individuals. The right axis is for type H inflows of job leavers (JL), and reentrants to the labor force (RE).
Source: Author's calculation.

Figure 11: Inflows of type H workers and less-educated individuals

individuals.

At the same time, the secular decline in inflows of less-educated workers mainly shows through to type H job leavers and reentrants to the labor force. Between the 1980s and 2010s, the monthly inflows of individuals whose educational attainment is high-school graduation or less decline by 1.2 million (black line in Figure 11). During the same period, the monthly inflows of type H job leavers and reentrants drop by 0.7 million, accounting for about 60% of the decline in the inflows of less-educated workers. This observation suggests that the downtrend in the inflows of type H job leavers and reentrants is closely related to the increased educational attainment of the labor force.[37]

Who do type H job leavers and reentrants represent? They are likely job switchers who experience a short intervening jobless spell, as their unemployment has a voluntary aspect and job switchers tend to have an intervening nonparticipation spell (Hall and Kudlyak (2019)). In other words, the joblessness of type H job leavers and reentrants is closely associated with churning in the labor markets.[38] Considering that one important symptom

[37]As the population becomes more highly educated, the number of newly unemployed workers with lower education has declined over time (Hornstein and Kudlyak (2019)).

[38]Lazear and Spletzer (2012) and Weingarden (2020) show that voluntary job leavers are the key piece of

of reduced dynamism is decreased churns (Decker et al. (2016), Decker et al. (2017)), the downtrend in their inflows suggests that the increased educational attainment of labor force is tied together with the decreased dynamism in the labor markets. As the workforce became more educated, job-specific human capital might have become more important, which likely lowered a worker's incentive to change jobs.

All told, the empirical analyses imply that the two related structural changes in the labor markets—the increased educational attainment of the labor force and the reduced dynamism—have particularly lowered the inflows of type H workers and shifted newly unemployed individuals toward those who tend to become long-term unemployed, which has gradually driven up the mean duration over the past decades.[39]

6.2 The source of limited recovery of type L unemployment-continuation probabilities of permanent job losers and reentrants to the labor force

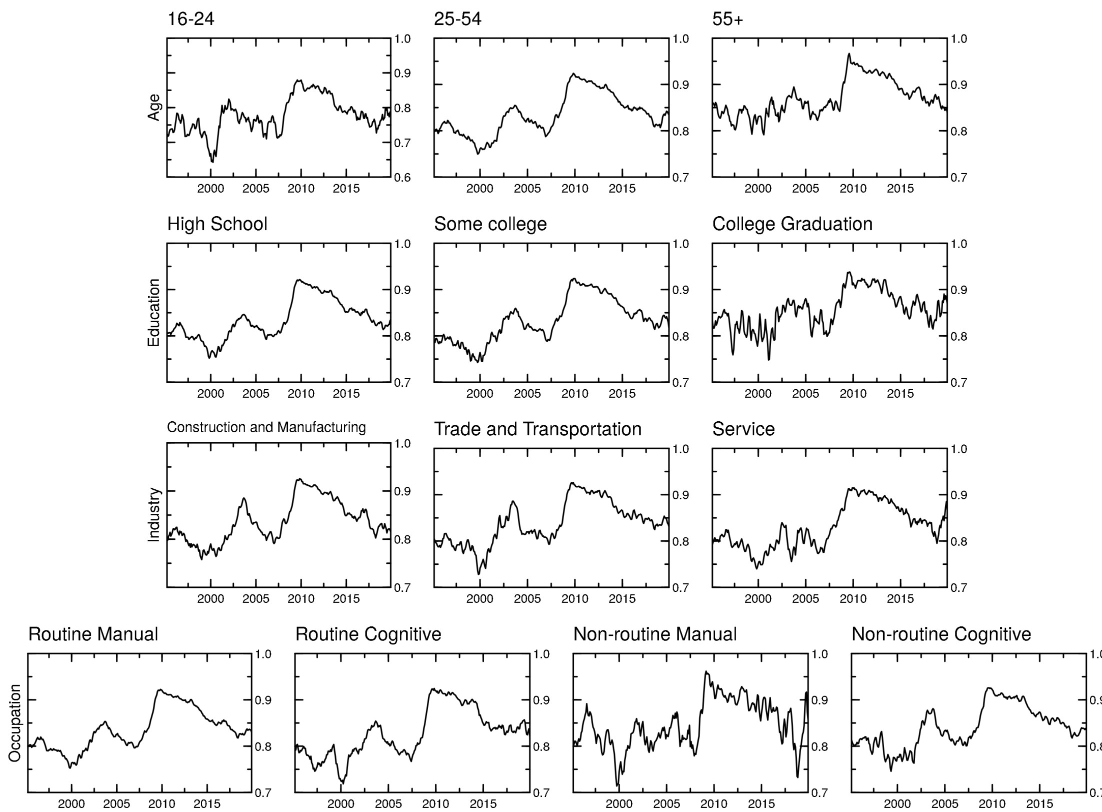
This section explores the source of the limited post-recession recovery of type L unemployment-continuation probabilities of permanent job losers and reentrants to the labor force from the 2000s. For this, I estimate the type L unemployment-continuation probability of permanent job losers and reentrants by gender, age, education, industry and occupation. The difficulty in estimating the nonlinear state space model with further disaggregate data is substantial due to large measurement errors caused by the disaggregation. Therefore, I take a non-parametric approach to approximate the type L probability, p^L_{jt}, from the following:

$$p^L_{jt} \approx \left(\frac{U^{7.+}_{jt}}{U^{4.+}_{j,t-3}}\right)^{\frac{1}{3}}. \tag{7}$$

churns in labor markets.

[39] There are studies such as Davis et al. (2010) and Weingarden (2017) that analyze the effects of reduced dynamism on unemployment. However, I do not acknowledge any research that discusses the effects of reduced dynamism on long-term unemployment or the mean duration of unemployment with unobserved worker heterogeneity taken into account.

The right-hand side of equation (7) is mainly determined by the unemployment-continuation probability of type L workers, because the majority of those unemployed for 4 months and over are type L workers in each group j.[40] I take a 12-month moving average to the estimates to smooth out the measurement errors.



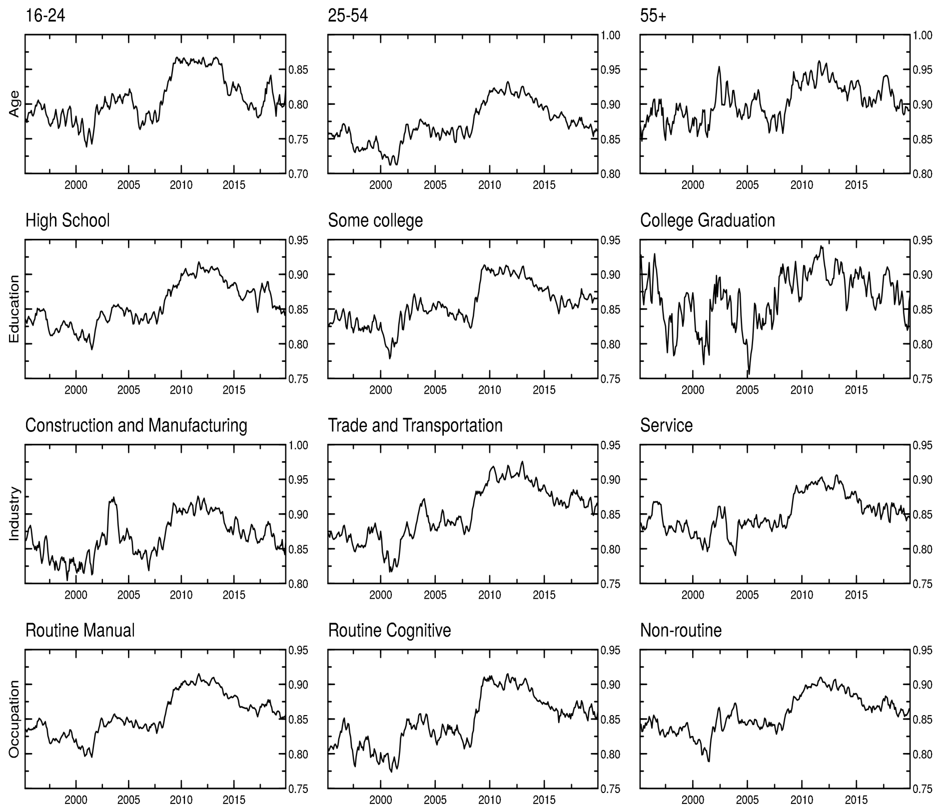
Source: Author's calculation.

Figure 12: Nonparametric estimates of p^L_{jt} among permanent job losers

Figures 12 and 13 report the estimates.[41] The pattern of limited recovery is broad-based, with the exceptions of college graduates and individuals aged 55 and over. In other words, it is unclear which observable worker characteristic is particularly associated with the limited

[40] GDD is not considered in this calculation. However, abstracting GDD does not influence the trend component in the estimates, as GDD mainly shifts the level of probability.

[41] Among reentrants to the labor force, I group workers with non-routine manual and non-routine cognitive occupations together due to the small number of observations of both groups.



Source: Author's calculation.

Figure 13: Nonparametric estimates of p^L_{jt} among reentrants to the labor force

recovery. Therefore, the worker attributes that hamper the unemployment duration of type L workers from recovering to pre-recession levels are ultimately unobserved in the CPS, though the factors have pervasive effects across various workers.

Meanwhile, recent studies (e.g., Macaluso (2019)) suggest that skills, though challenging to measure, might be an important determinant of labor market outcomes among job losers. Indeed, the close association between permanent job losers and type L workers also implies that skills might be the key unobserved factor. Workers who have skills not demanded by firms may lose their jobs permanently and stay unemployed for a longer period of time. Considering some reentrants are permanent job losers who left the labor force and came back, type L reentrants are also likely to be those who experience skill mismatch. In this context, the uptrend in unemployment-continuation probabilities among type L permanent job losers and reentrants might be driven by skill-biased technological changes. To verify this hypothesis, however, one needs new data sources.

7 Conclusion

This paper demonstrates that a statistical model can capture the existence of unobserved worker heterogeneity and its consequences for changes in the aggregate duration of unemployment. Unobserved heterogeneity is important in not only cyclical but also low frequency variations in the aggregate duration of unemployment. The close link between permanent job loss and type L unemployment implies the changing demand for skills might be an important factor that makes a worker stay unemployed for a long time. In this context, it is crucial to think about the distribution of unobserved heterogeneity as a dynamic process. This research also suggests new data sources are needed to better identify the group of workers that are critical in understanding long-term unemployment.

References

Acemoglu, Daron, "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, March 2002, *40* (1), 7–72.

Ahn, Hie Joo and James D. Hamilton, "Heterogeneity and Unemployment Dynamics," *Journal of Business & Economic Statistics*, 2020, *38* (3), 554–569.

— **and** — , "Measuring Labor-Force Participation and the Incidence and Duration of Unemployment," *Review of Economic Dynamics*, Forthcoming.

Baker, Michael, "Unemployment Duration: Compositional Effects and Cyclical Variability," *American Economic Review*, March 1992, *82* (1), 313–321.

Darby, Michael R., John C. Haltiwanger, and Mark W. Plant, "The Ins and Outs of Unemployment: The Ins Win," NBER Working Papers 1997, National Bureau of Economic Research, Inc August 1986.

Davis, Steven J., R. Jason Faberman, John Haltiwanger, Ron Jarmin, and Javier Miranda, "Business Volatility, Job Destruction, and Unemployment," *American Economic Journal: Macroeconomics*, April 2010, *2* (2), 259–287.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda, "Declining Business Dynamism: What We Know and the Way Forward," *American Economic Review*, May 2016, *106* (5), 203–07.

_ , _ , _ , and _ , "Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown," *American Economic Review*, May 2017, *107* (5), 322–326.

Farber, Henry S., Jesse Rothstein, and Robert G. Valletta, "The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012-2013 Phase-Out," *American Economic Review*, May 2015, *105* (5), 171–176.

Fujita, Shigeru and Giuseppe Moscarini, "Recall and Unemployment," *American Economic Review*, December 2017, *107* (12), 3875–3916.

Gregory, Victoria, Guido Menzio, and David G. Wiczer, "The Alpha Beta Gamma of the Labor Market," NBER Working Papers 28663, National Bureau of Economic Research, Inc April 2021.

Hall, Robert E. and Marianna Kudlyak, "Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics," NBER Working Papers 25625, National Bureau of Economic Research, Inc February 2019.

_ and Sam Schulhofer-Wohl, "Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers," *American Economic Journal: Macroeconomics*, January 2018, *10* (1), 1–32.

Hornstein, Andreas, "Accounting for unemployment: the long and short of it," Technical Report 2012.

_ and Marianna Kudlyak, "Aggregate Labor Force Participation and Unemployment and Demographic Trends," February 2019, (2019-7).

Jarosch, Gregor and Laura Pilossoph, "Statistical Discrimination and Duration Dependence in the Job Finding Rate," *The Review of Economic Studies*, 09 2018, *86* (4), 1631–1665.

Katz, Lawrence F. and Bruce D. Meyer, "The impact of the potential duration of unemployment benefits on the duration of unemployment," *Journal of Public Economics*, February 1990, *41* (1), 45–72.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 2013, *128* (3), 1123–1167.

— , — , — , and Lawrence F. Katz, "Long-Term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Nonparticipation," *Journal of Labor Economics*, 2016, *34* (S1), 7–54.

Krueger, Alan B., Judd Cramer, and David Cho, "Are the Long-Term Unemployed on the Margins of the Labor Market?," *Brookings Papers on Economic Activity*, 2014, *45* (1 (Spring), 229–299.

Lazear, Edward P. and James R. Spletzer, "Hiring, Churn, and the Business Cycle," *American Economic Review*, May 2012, *102* (3), 575–579.

Macaluso, Claudia, "Skill Remoteness and Post-Layoff Labor Market Outcomes," 2019.

Polivka, Anne E. and Stephen M. Miller, *The CPS after the Redesign: Refocusing the Economic Lens*, University of Chicago Press, January

Shimer, Robert, Katarina Borovickova, and Fernando Alvarez, "The Proportional Hazard Model: Estimation and Testing using Price Change and Labor Market Data," Technical Report 2015.

Weingarden, Alison, "Job Reallocation and Unemployment in Equilibrium," FEDS Notes, Washington: Board of Governors of the Federal Reserve System April 2017.

— , "Worker Churn at Establishments over the Business Cycle," FEDS Notes, Board of Governors of the Federal Reserve System August 2020.