Heterogeneity and Unemployment Dynamics[*]

Hie Joo Ahn[†]and James D. Hamilton[‡]

May 1, 2014
Revised: September 1, 2018

Abstract

Many previous papers have studied the contribution of inflows and outflows to the cyclical variation in unemployment, but ignored the critical role of unobserved heterogeneity across workers. This paper develops new estimates of unemployment inflows and outflows that allow for unobserved heterogeneity as well as direct effects of unemployment duration on unemployment-exit probabilities. With this approach we can measure the contribution of different shocks to the short-run, medium-run, and long-run variance of unemployment as well as to specific historical episodes. We conclude that changes in the composition of new inflows into unemployment are the most important factor in economic recessions.

Keywords: business cycles, Great Recession, unemployment duration, unobserved heterogeneity, duration dependence, state space model, extended Kalman filter

Introduction

What accounts for the sharp spike in the unemployment rate during recessions? The answer traditionally given by macroeconomists was that falling product demand leads firms to lay off workers, with these inflows into unemployment a key driver of economic downturns. That view has been challenged by Hall (2005), Shimer (2012) and Hall and Schulhofer-Wohl (2017), who argued that cyclical fluctuations in the unemployment rate are instead primarily driven by declines in the job-finding rates for unemployed workers. By contrast, Yashiv (2007), Elsby, Michaels and Solon (2009), Fujita and Ramey (2009), and Fujita (2011) concluded that flows into unemployment are as important or more important than outflows as cyclical drivers of the unemployment rate.

One factor that has been missing in this debate is the role of unobserved heterogeneity. When there are differences across workers, changes in inflows into unemployment change the composition of the pool of unemployed and thereby change the outflow rate. Suppose for example that 80% of the newly unemployed (whom we call type H) have unemployment-continuation probability of 35% and 20% (type L) have probability of 85%. If we look at how many of those individuals are still unemployed n months later, type L make up a larger fraction of the remaining unemployed. This will cause the observed average continuation probability of the group to rise as a function of n (see Figure 1). If in the current month there is an increase in the number of type L job losers, they are likely to stay unemployed longer and will bring down the average job-finding probability in future months.

This paper develops a full dynamic model of the interaction between unemployment inflows, outflows, and unobserved heterogeneity. Developing a complete statistical model allows us to measure formally the fraction of the error in forecasting unemployment at any horizon that is attributable to inflows and outflows. Our framework further allows us to decompose the forecast error associated with any given historical episode into the respective contributions of inflows and outflows. All this is new to this literature.

Darby, Haltiwanger and Plant (1986) concluded that changes in the composition of inflows are indeed the cause of changes in future outflow rates, a hypothesis that Baker (2012) and others have come to refer to as the "heterogeneity hypothesis." A large literature has examined the heterogeneity hypothesis. But all of the previous papers in this literature posed the question in

terms of differences in observable characteristics.[1]

Why is unobserved heterogeneity important? Consider this striking feature of the data: on average, someone who is newly unemployed has less than a 50% chance of still being unemployed next month. By contrast, someone who has been unemployed for 4 months or longer has an 80% chance of remaining unemployed next month.[2] The newly unemployed during the Great Recession had better job-finding prospects than did the long-term unemployed during the strongest economic boom. These huge differences in unemployment-continuation probabilities remain regardless of the observable characteristics on which one may try to condition.[3] Any pool of unemployed individuals who share any given observed characteristics is going to become increasingly represented by those within that group who have higher ex ante continuation probabilities the longer the period of time for which the individuals have been unemployed.

In addition to unobserved heterogeneity, the high unemployment-continuation probabilities of the long-term unemployed could arise if the experience of being unemployed for a longer period of time directly changes the employment probability for a fixed individual, an effect referred to as "genuine duration dependence." A large literature has discussed the difficulty of distinguishing genuine duration dependence from unobserved heterogeneity.[4] A common approach has been to assume that there is no variation over time in unobserved heterogeneity, in which case identification can be achieved by observing repeated spells of unemployment for a given individual (Honoré,1993).

However, unobserved heterogeneity arises in part from factors such as specific skill sets. The demand for these varies over time with changes in technology and business conditions. Modeling this requires allowing the distribution characterizing unobserved heterogeneity to be time varying. Our paper uses a proportional hazards specification in which the identifying assumptions are that genuine duration dependence does not change over time while the distribution characterizing un-

---

[1] Baker (1992), Shimer (2012), and Kroft, et al. (2016) found that observed variables contributed little to variation over time in long-term unemployment rates, while Aaronson, Mazumder and Schechter (2010), Bachmann and Sinning (2012), Barnichon and Figura (2015), Hall (2014), and Hall and Schulhofer-Wohl (2017) documented important differences across observable characteristics. Elsby, Michaels and Solon (2009) found that incorporating observable heterogeneity reduced the imputed role of cyclical variation in unemployment exit rates.

[2] Let $\tilde{U}_t^{n,+}$ denote the seasonally unadjusted number of individuals in month t who report having been unemployed for n months or longer at that time. The seasonally unadjusted monthly unemployment-continuation probability for the long-term unemployed was calculated as $\tilde{p}_t^{4,+} = (\tilde{U}_{t+3}^{7,+}/\tilde{U}_t^{4,+})^{1/3}$. The probability for the newly unemployed was calculated as the solution to $\tilde{p}_t^1(1 + \tilde{p}_t^1) = \tilde{U}_{t+1}^{2,3}/\tilde{U}_t^1$.

[3] For example, for individuals who gave involuntary permanent separation as the reason for unemployment, the average unemployment-continuation probability since 1994:M1 (the sample for which this finer separation exists) was 70% for the newly unemployed and 84% for the long-term unemployed.

[4] See for example Heckman and Singer (1984a,b), Honoré (1993), and van den Berg (2001).

observed heterogeneity evolves over time according to a simple process. Our paper is the first to describe cyclical changes in unobserved heterogeneity and to analyze their importance for unemployment dynamics. Although our approach relies on some parametric assumptions, we will show that it provides a natural, compelling, and robust way of interpreting the observed data.

Section 1 introduces the data that we will use in this analysis based on the number of job-seekers each month who report they have been looking for work at various search durations. We describe the accounting identities that will later be used in our full dynamic model and use average values of observable variables over the sample to explain the intuition behind our main results.

In Section 2 we extend this framework into a full dynamic model in which we represent heterogeneity in terms of two different types of workers at any given date. Type H workers have a higher ex ante probability of exiting unemployment than type L workers. Our model postulates that for each type, the number of newly unemployed individuals as well as the probability of exiting unemployment at each date evolve according to unobserved random walks. We also allow for nonmonotonic time-invariant genuine duration dependence. We show how to approximate the likelihood function for the observed unemployment data and form an inference about the state variables at every date using an extended Kalman filter.[5]

Empirical results are reported in Section 3. We find that variation over time in the inflows of the newly unemployed are more important than outflows in accounting for errors in predicting aggregate unemployment at all horizons. Inflow and outflow probabilities for type L workers are more important than those for type H workers, and account for 90% of the uncertainty in predicting unemployment 2 years ahead. In recessions since 1990, shocks to the inflows of type L workers were the most important cause of rising unemployment during the recession.

Section 4 provides corroborating evidence based on reduced-form VARs. We first use a bivariate VAR for inflows and outflows to show that a great deal of the observed variation in outflows could have been predicted on the basis of earlier values of inflows. We further show that changes in the composition of inflows have additional predictive power for future outflows, and demonstrate that

---

[5] Our approach is closely related to that in Hornstein (2012), who used dynamic accounting identities to interpret aggregate panel dynamics in a similar way to that in our paper. However, Hornstein's model was unidentified– in terms of the discussion of identification in Section 1, his model has 5 unknowns and only 4 equations. As a result, his specification did not allow him to calculate the likelihood function for the observed data or forecasts of unemployment or duration. By contrast, our model generates values for all these along with the optimal statistical inference about the various shocks driving the observed dynamics of unemployment.

changes in the level and composition of inflows account for most of the increase in unemployment during the Great Recession.

In Section 5 we investigate the robustness of our approach to various alternative specifications, including alternative methods to account for the change in the CPS questionnaire in 1994, allowing for correlation between the innovations of the underlying structural shocks in our model, and the possible effects of time aggregation. While such factors could produce changes in some of the details of our inference, our overall conclusions (summarized in Section 6) appear to be quite robust.

1   Observable implications of unobserved heterogeneity

The purpose of this section is to use steady-state calculations to explain how our approach allows for both unobserved heterogeneity and genuine duration dependence and provide the intuition behind some of the results that will be found in Section 3 using our full dynamic model.

The Bureau of Labor Statistics reports for each month t the number of working-age individuals who have been unemployed for less than 5 weeks. Our baseline model is specified at the monthly frequency, leading us to use the notation U_t^1 for the above BLS-reported magnitude, indicating these individuals have been unemployed for 1 month or less as of month t. BLS also reports the number who have been unemployed for between 5 and 14 weeks (or 2-3 months, denoted $U_t^{2.3}$), 15-26 weeks ($U_t^{4.6}$) and longer than 26 weeks ($U_t^{7.+}$). One reason the BLS reports the data in terms of these duration aggregates is to try to minimize the role of measurement error by averaging within broad groups. We will do the same. Our theoretical model will generate a prediction of the number of unemployed at every monthly duration, but we will only use the model's implications about broad duration aggregates for purposes of calculating the likelihood function of the observed data.[6] Notwithstanding, when reporting on long-term unemployment, many BLS publications[7] further break down the $U_t^{7.+}$ category into those unemployed with duration 7-12 months ($U_t^{7.12}$) and those with duration longer than 1 year ($U_t^{13.+}$). Since long-term unemployment is also a major interest in our investigation, we have used the raw CPS micro data from which the usual publicly

---

[6]In January 2011 the BLS changed the maximum allowable unemployment duration response from 2 years to 5 years. Although this affected the BLS's own estimate of average duration of unemployment, it did not change the total numbers unemployed by the duration categories we use. This is another reason to favor our approach, which relies only on aggregated data.

[7]See for example Ilg and Theodossiou (2012).

reported aggregates are constructed to create these last two series for our study.[8]

The average values over the full sample of these five observed variables on which our inference will be based are reported in the first row of Table 1. Our focus is on the following question– of those individuals who are newly unemployed at time t, what fraction will still be unemployed at time $t + k$? We presume that the answer to this question depends not just on aggregate economic conditions over the interval $(t, t + k)$ but also on the particular characteristics of those individuals. Let w_{it} denote the number of people of type i who are newly unemployed at time t. Thus

$$U_t^1 = \sum_{i=1}^{I} w_{it}. \tag{1}$$

We define $P_{it}(k)$ as the fraction of individuals of type i who were newly unemployed in $t - k$ and are still unemployed at t. The total number of individuals who have been unemployed for exactly $k + 1$ months at time t is given by

$$U_t^{k+1} = \sum_{i=1}^{I} w_{i,t-k} P_{it}(k). \tag{2}$$

We first examine what we could infer about unobserved types based only on the historical average values $\bar{U}^1, \bar{U}^{2.3}, \bar{U}^{4.6}, \bar{U}^{7.12}$, and $\bar{U}^{13.+}$, and then will consider what additional information can be learned from variation over time in these five variables.

1.1   Inference using historical average values alone

Suppose for purposes of this section only that the number of newly unemployed individuals of each type remained constant over time at values w_i and also that the probabilities that individuals of each type remain unemployed in any given month are constants p_i for $i = 1, ..., I$. Consider first the case when there is only one type of worker ($I = 1$). Under these assumptions (2) would simplify to $U^{k+1} = wp^k$. Given the average observed values for U^k for two different values of k, we could then estimate the values of w and p, for example, $\hat{w} = \bar{U}^1$ and $\hat{p} = \bar{U}^2/\bar{U}^1$. As noted above, we regard aggregate measures like $U_t^{2.3}$ as more reliable than a specific estimate such as U_t^2 that could be constructed from CPS micro data, and therefore use instead $\hat{p} + \hat{p}^2 = \bar{U}^{2.3}/\bar{U}^1$. The

---

[8]See Appendix A for further details of data construction.

estimated values for $\hat{w}$ and $\hat{p}$ that result from this equation are reported in row 2 of Table 1 and plotted in Panel A of Figure 2. Most of the newly unemployed find jobs quickly ($\hat{p} = 0.48$). But if workers who had been unemployed for more than 3 months also had this same job-finding rate, there would be far fewer workers in the 4-6 month, 7-12, and 13+ categories than we observe in the data, as represented by the black circles in Figure 2.[9]

Consider next the case when there are $I = 2$ types of workers. In this case (2) becomes

$$U^{k+1} = w_L p_L^k + w_H p_H^k. \tag{3}$$

This equation describes the average number of individuals who have been unemployed for $k + 1$ months as the sum of two different functions of k, with each of the two functions being fully characterized by two parameters (w_i and p_i). The solid red curve in Panel B of Figure 2 plots the first function ($w_L p_L^k$), while the dotted blue curve plots the sum. Given observed values of $\bar{U}^1, \bar{U}^{2.3}, \bar{U}^{4.6}$, and $\bar{U}^{7.12}$, we could estimate the four parameters (w_L, w_H, p_L, p_H) to match exactly those four observations, as in Panel B of Figure 2 and row 3 of Table 1.[10] These estimates imply that type H individuals comprise 78% of the initial pool of unemployed U^1. But the unemployment-continuation probability for type H individuals ($p_H = 0.36$) is much lower than for type L ($p_L = 0.85$). Because the type H are likely to find jobs relatively quickly, there are very few type H individuals included in U^n for durations n beyond 4 months, as seen in Panel B of Figure 2. The key feature of the observed data (represented by the black dots in Figure 2) that gives rise to this conclusion is the fact that the numbers drop off very quickly at low durations (as most of the type H workers find jobs), but after that much more slowly (as the remaining type L workers continue searching).

What about when $I > 2$? In this case we can still get a useful characterization of heterogeneity across workers by separating them into two broad types. Specifically, for any true values for w_i and p_i for $i = 1, ..., I > 2$ and any observed 4 durations k_1, k_2, k_3, k_4, we can find values for the 4

---

[9] The black circles are used as a visual device to summarize the interval averages from row 1 of Table 1. Specifically, they are the implied values at the particular durations 1, 3, 5, 9.5, and 15 months from a flexible functional form (equation (6)) that could predict numbers for every duration and whose predictions exactly match the observed average values for the five observations in row 1 of Table 1.

[10] Specifically, the four functions are obtained from equations (10)-(13) below for the special case when the left-hand variables represent historical averages and on the right-hand side we set $w_{it} = w_i$, $P_{it}(k) = p_i^k$, and $r_t^x = 0$.

parameters $(\hat{w}_L, \hat{w}_H, \hat{p}_L, \hat{p}_H)$ to approximate the cross-sectional distribution as the solutions to

$$\hat{w}_L \hat{p}_L^k + \hat{w}_H \hat{p}_H^k = \sum_{i=1}^{I} w_i p_i^k \quad \text{for } k = k_1, k_2, k_3, k_4. \tag{4}$$

Note that if we only observed 4 duration categories, a mixture of two types is a fully general characterization of heterogeneity in the sense that it can completely describe all the features observable in the data and provides the identical fit to the observed data as would a specification with $I > 2$.[11] Given measurement error in the CPS data, we do not believe we can reliably use more than 5 observed duration categories, meaning estimation of more than $I = 2$ types is infeasible using these data. In other data sets and in somewhat different settings from ours, Ham and Rea (1987), Van den Berg and van Ours (1996), and Van den Berg and van der Klaauw (2001) tested for the number of types and found $I = 2$ is sufficient to capture heterogeneity in the data sets they analyzed. In this paper we will represent heterogeneity in terms of a mixture of two types, though we view this primarily as a convenient approximation.[12]

Although we did not use the fifth data point, $\bar{U}^{13.+}$, in estimating these parameters, the framework generates a prediction for what that observation would be.[13] This is reported in the last entry of row 3 of Table 1 to be 621,000 which is quite close to the observed value of 664,000. The feature of the data that produced this result is that the observed numbers fall off at close to a constant exponential rate once we get beyond 4 months, as the simple mixture model would predict.

Alternatively, we could equally well describe the observed averages using a model in which there is only genuine duration dependence (GDD). Suppose that an individual who has been unemployed for τ months has a probability $p(\tau)$ of still being unemployed the following month. We can always write this in the form $p(\tau) = \exp(-\exp(d_\tau))$ for d_τ an arbitrary function of τ. For example, we could fit the 5 observations in the first row of Table 1 perfectly if we used $w = \bar{U}^1$ along with a

---

[11] This result can be viewed as an illustration of Theorem 3.1 in Heckman and Singer (1984a).

[12] If in a true population for which $I > 2$ there is an increase in the level of inflows with no change in outflows or relative composition ($w_i^* = \lambda w_i$ and $p_i^* = p_i$ for $i = 1, ..., I$), then a 2-mixture approximation would correctly conclude that only the level of inflows has changed with no change in composition or in outflows; that is, equation (4) would have the solution $\hat{w}_i^* = \lambda \hat{w}_i$ and $\hat{p}_i^* = \hat{p}_i$ for $i = H, L$. Likewise if there is a proportional change in all outflow probabilities with no change in composition or inflows in a population mixture of I types ($w_i^* = w_i$ and $p_i^* = \lambda p_i$ for $i = 1, ..., I$), the 2-type approximation (4) would correctly conclude that $\hat{w}_i^* = \hat{w}_i$ and $\hat{p}_i^* = \lambda \hat{p}_i$ for $i = H, L$ since for each k the left and right sides of the equation are then multiplied by λ^k.

[13] Following Hornstein (2012) we truncate all calculations at 48 months in equation (14). Most of the models considered in this paper imply essentially zero probability of an unemployment spell exceeding 4 years in duration.

4-parameter representation for d_τ such as[14]

$$d_\tau = \delta_0 + \delta_1\tau + \delta_2\tau^2 + \delta_3\tau^3. \tag{5}$$

A large number of empirical studies have assumed Weibull durations, essentially corresponding to $\delta_2 = \delta_3 = 0$. The values for δ_j that would exactly fit the historical averages are reported in row 4 of Table 1 and the implied function $p(\tau)$ is plotted in panel A of Figure 3. Note that in contrast to the popular Weibull assumption and most theoretical models, the fitted function (5) is not monotonic.

If we were willing to restrict the functional form of GDD to the Weibull case, we could also interpret the historical averages as resulting from a combination of unobserved heterogeneity and GDD. Suppose we assumed proportional hazards[15] and represent the probability that an individual of type i who has been unemployed for τ months will still be unemployed the following month as

$$p_i(\tau) = \exp\{-\exp[x_i + d_\tau]\} \tag{6}$$

with implied unemployment counts

$$U^{k+1} = \sum_{i=L,H} w_i p_i(1) p_i(2) \cdots p_i(k). \tag{7}$$

The value of x_i for $i = H, L$ reflects cross-sectional heterogeneity in unemployment-continuation probabilities and d_τ captures genuine duration dependence. As noted by Katz and Meyer (1990), this double-exponential functional form is a convenient way to implement a proportional hazards specification so as to guarantee a positive hazard[16], a feature that will be very helpful for the generalization in the following section in which we will allow for variation of x_{it} over time. Sup-

---

[14] Specifically, we calculate $U^{k+1} = wp(1)p(2)\cdots p(k)$ and find the values of $w, \delta_0, \delta_1, \delta_2, \delta_3$ to match the observed values in row 1 of Table 1.

[15] Alvarez, Borovičková, and Shimer (2016) concluded that proportional hazards is not consistent with the observed data. Their identifying assumption was that the heterogeneous characteristics of individual i do not change even if the individual is observed in different decades. But employers' demands for the specific skills of individual i surely change over time. For example, the demand for carpenters varies over housing booms and busts. By contrast, our specification in the following section allows both an individual's identification with a particular group as well as the group's average unemployment-continuation probabilities to be continually changing, an approach that gives a proportional-hazards specification considerably more flexibility.

[16] Consider an individual i who has been unemployed for τ months as of the beginning of month t and let the hazard within month t be $\lambda_{i,t,\tau} = \exp(x_{it})\exp(d_\tau)$ where the exponentiation is a device to guarantee that the hazard is positive for any x_{it} and d_τ. The meaning of the hazard is that if we divide month t into n subintervals, the probability that individual i exits unemployment in the interval $(s, s + 1/n)$ is $\lambda_{i,t,\tau}/n + o(1/n)$ from which the

pose we were willing to model GDD using a one-parameter function, say $d_\tau = \delta(\tau - 1)$. Then we could find a value for the 5 parameters $w_L, w_H, x_L, x_H, \delta$ so as to fit the 5 time-series averages $\bar{U}^1, \bar{U}^{2.3}, \bar{U}^{4.6}, \bar{U}^{7.12}$, and $\bar{U}^{13.+}$ exactly. These values are reported in row 5 of Table 1. The implied value for δ is close to zero, and the other parameters are close to those for the pure cross-sectional heterogeneity specification of row 3. Thus for this particular parametric example, we would conclude that cross-sectional heterogeneity is much more important than genuine duration dependence in accounting for why observed unemployment-continuation probabilities rise with duration of unemployment. The feature of the data that gave rise to this conclusion is that the 4-parameter pure heterogeneity model gives a very good prediction of all five observations.

1.2   Inference using changes over time

Next consider what we can discover using time-series variation in the observed aggregates. Suppose we repeat the above exercises only using data around the Great Recession. Row 7 of Table 1 and Panel C of Figure 2 show the results if we tried to explain these numbers entirely in terms of unobserved heterogeneity. The implied value for the unemployment-continuation probability for type L individuals, $p_L = 0.89$, is only slightly higher than the value 0.85 fit to the full historical sample. The reason is that the function $\bar{U}^n$ drops off after $n = 4$ months at only a slightly slower rate than it did historically. However, we would infer that the inflow of new type L individuals, $w_L = 1,065$ is much higher than the historical average value of 690, in order to account for the fact that $\bar{U}^n$ is now dropping off after 4 months from a much higher base. We again find that the 4-parameter model does a reasonable job of anticipating the fifth unused data point.

If we instead tried to explain the recent averages purely in terms of GDD, we would use the parameter values from row 8 of Table 1. These again could fit the data perfectly, albeit relying on a function with odd oscillations (see panel B of Figure 3). Although it is mathematically possible to describe the data with this equation, it would be difficult to motivate a theory of why GDD should have changed shape in this way. It requires for example a steeper initial slope to the curve in panel B of Figure 3 when economic conditions worsened, corresponding to the claim that the scarring

---

probability that the individual is still unemployed at the beginning of month $t + 1$ is

$$\lim_{n \to \infty} [1 - \lambda_{i,t,\tau}/n + o(1/n)]^n = \exp(-\lambda_{i,t,\tau}) = \exp[-\exp(x_{it})\exp(d_\tau)].$$

associated with unemployment is more severe during a recession. But this is directly contradicted by the experimental finding of Kroft, Lange, and Notowidigdo (2013) that potential employers pay less attention to applicants' duration of unemployment when the labor market is weaker.

These concerns notwithstanding, it is possible to allow for both an unrestricted nonmonotonic functional form for GDD as well as unobserved heterogeneity once we take account of changes over time. Suppose for example we were to pool the observations from the first row of Table 1 (the full-sample averages) together with those in row 6 (behavior around the Great Recession), giving us a total of 10 observations. If we took the view that the unobserved heterogeneity parameters may have changed over the cycle but that the GDD function d_τ in (6) is time-invariant, we would then be able to generalize d_τ to be a function of τ determined by two parameters, say δ_1 and δ_2, and use the ten observations to infer ten unknowns (values of w_H, w_L, x_H, x_L for the two subsamples along with the parameters δ_1 and δ_2). Generalizing a little further, if we use observations across 4 different subsamples we could infer values of w_H, w_L, x_H, x_L for each subsample along with a completely unrestricted nonmonotonic GDD function as in (5). In fact, if we were able to use all five observations on $U_t^1, U_t^{2.3}, U_t^{4.6}, U_t^{7.12}, U_t^{13.+}$ for every date t, we could even allow for some modest variation over time in the GDD function $d_{\tau t}$, and indeed such a specification will be included in the general results reported in Section 5.

We have used steady-state calculations in this section primarily to explain the intuition for where the identification is coming from. Nevertheless, it turns out that the key conclusions of the above steady-state calculations– that the majority of newly unemployed individuals can be described as type H who find jobs quickly, that dynamic sorting based on unobserved heterogeneity appears to be much more important than genuine duration dependence in explaining why a longer-term unemployed individual is less likely to exit unemployment, and that the key driver of economic recessions is an increased inflow of newly unemployed type L individuals– will also turn out to characterize what we will find as we now turn to a richer dynamic model.

2 Dynamic formulation

Our dynamic model is a generalization of (6) in which outflow probabilities for each type of individual change over time. We assume that for type i workers who have already been unemployed

for τ months as of time $t-1$, the fraction who will still be unemployed at t is given by

$$p_{it}(\tau) = \exp[-\exp(x_{it} + d_\tau)] \quad \text{for} \quad \tau = 1, 2, 3, ... \tag{8}$$

where d_τ is a third-order polynomial as in equation (5).[17] We also allowed inflows for each type to vary over time, letting w_{it} change each month. Note the identifying assumption is that the contribution of genuine duration dependence d_τ, while of the completely general functional form used in Figure 3, does not vary over time.[18] We now specify a state-space model where the dynamic behavior of the observed vector $y_t = (U_t^1, U_t^{2.3}, U_t^{4.6}, U_t^{7.12}, U_t^{13.+})'$ is determined as a nonlinear function of latent dynamic variables– the inflows and outflow probabilities for unemployed individuals with unobserved heterogeneity. Due to the nonlinear nature of the resulting model, we draw inference on the latent variables using the extended Kalman filter.

2.1 State-space representation

Our baseline model assumes that the elements of $\xi_t = (w_{Ht}, w_{Lt}, x_{Ht}, x_{Lt})'$ each evolve as random walks, e.g.,

$$w_{Ht} = w_{H,t-1} + \varepsilon_{Ht}^w. \tag{9}$$

A random walk is the typical assumption in dynamic latent-variable or time-varying-parameter models and has proven to be a flexible and parsimonious way to adapt inference to a variety of sources of changing conditions or possible structural breaks.[19] Note also that equation (9) is an unambiguous improvement over the steady-state calculations described in the previous section (and invoked in the majority of previous studies in this literature), and includes the steady-state formulation as a special case when the variance of ε_{Ht}^w is zero. We have also experimented with a model in which we assume AR(1) dynamics for the latent variables with autoregressive coefficients

---

[17]We found that the numerical search to find the maximum likelihood estimates performed best when we expressed this function in terms of scaled Chebyshev polynomials:

$$d_\tau = \tilde{\delta}_1((\tau-1)/48) + \tilde{\delta}_2[2((\tau-1)/48)^2 - 1] + \tilde{\delta}_3[4((\tau-1)/48)^3 - 3((\tau-1)/48)].$$

[18]In fact our approach can also allow for modest time variation. In the robustness analysis in Section 5 we replace d_τ with $d_{t\tau}$ which changes with t in a restricted way.

[19]See for example Baumeister and Peersman (2013).

estimated by maximum likelihood. We found the coefficient estimates to be very close to unity and the resulting inference very similar to those for our baseline random walk specification.

The intuition for how the extended Kalman filter works is as follows. We will have formed an inference about the value of ξ_t based on the data we observed through date t. For example, we could use the steady-state calculations of Section 2 on a small initial sample of observed $y_1, .., y_{t_0}$ to form an initial inference about $w_{H,t_0}, w_{L,t_0}, x_{H,t_0}, x_{L,t_0}$, which would imply values for $U_{t_0}^n$ for every n from equation (3) based on the average values for that initial sample.[20] A random walk means that we enter period $t + 1$ initially expecting it to look like t. This would imply predicted values for the five variables observed at $t + 1$. If $U_{t+1}^{13.+}$ is higher than predicted, it would be an indication that p_L has gone up (since there are essentially no type H individuals included in $U_t^{12.+}$). If $U_{t+1}^{2.3}$ is higher than predicted even with this higher value for $p_{L,t+1}$ it means that $p_{H,t+1}$ has likely gone up as well. If U_{t+1}^1 is higher than U_t^1, we know that either w_L or w_H must have gone up. Given the 5 new observations in y_{t+1}, we have more than enough information to update an inference about all 4 elements of ξ_{t+1}. Proceeding sequentially through the observed sample in this way, we can form an inference about ξ_t for every date and at the same time improve our inference about the previous history. The final revised inference about the state at date t based on seeing the full sample of data through date T is referred to as the smoothed inference, denoted $\hat{\xi}_{t|T}$.

Another key detail of our approach is that we allow for the possibility that unemployment counts are all contaminated by error. The durations in CPS are in part self-reported and respondents make a variety of errors. We assume that each element of y_t has an associated measurement error $r_t = (r_t^1, r_t^{2.3}, r_t^{4.6}, r_t^{7.12}, r_t^{13.+})'$. Our identification assumption is that the measurement error is white noise, meaning that the inference is only adjusted for changes in the observed variables that prove to be persistent. The observation equations can then be written as follows,

$$U_t^1 = \sum_{i=H,L} w_{it} + r_t^1 \tag{10}$$

$$U_t^{2.3} = \sum_{i=H,L} [w_{i,t-1} P_{it}(1) + w_{i,t-2} P_{it}(2)] + r_t^{2.3} \tag{11}$$

---

[20] Our estimates below start with $t_0 = 1976$:M1 and set $\hat{\xi}_{t_0|t_0}$ to the solution to the steady-state model over the period 1972:M1-1976:M1. Our approach allows the true value ξ_{t_0} to differ from this estimate with a very large variance, so that the initial estimate has a very limited contribution. See Appendix B for details.

$$U_t^{4.6} = \sum_{i=H,L} \sum_{k=3}^{5} [w_{i,t-k} P_{it}(k)] + r_t^{4.6} \tag{12}$$

$$U_t^{7.12} = \sum_{i=H,L} \sum_{k=6}^{11} [w_{i,t-k} P_{it}(k)] + r_t^{7.12} \tag{13}$$

$$U_t^{13.+} = \sum_{i=H,L} \sum_{k=12}^{47} [w_{i,t-k} P_{it}(k)] + r_t^{13.+} \tag{14}$$

$$P_{it}(j) = p_{i,t-j+1}(1) p_{i,t-j+2}(2) ... p_{it}(j). \tag{15}$$

We can arrive at the likelihood function for the observed data $\{y_1, ..., y_T\}$ by assuming that the measurement errors are independent Normal,[21] $r_t \sim N(0, R)$, with $R = \text{diag}(R_1^2, R_{2.3}^2, R_{4.6}^2, R_{7.12}^2, R_{13.+}^2)$ whose diagonals are the variances of $r_t^1, r_t^{2.3}, r_t^{4.6}, r_t^{7.12}$ and $r_t^{13.+}$ respectively.

Let ξ_t be the vector $(w_{Lt}, w_{Ht}, x_{Lt}, x_{Ht})'$ and $\varepsilon_t = (\varepsilon_{Lt}^w, \varepsilon_{Ht}^w, \varepsilon_{Lt}^x, \varepsilon_{Ht}^{x\prime})'$. Our assumption that the latent factors evolve as random walks would be written as

$$\underbrace{\xi_t}_{4 \times 1} = \xi_{t-1} + \underbrace{\varepsilon_t}_{4 \times 1} \tag{16}$$

for ε_t a (4×1) vector with variance matrix $\Sigma = \text{diag}((\sigma_L^w)^2, (\sigma_H^w)^2, (\sigma_L^x)^2, (\sigma_H^x)^2)$. In Section 5 we will also report results for a specification in which the shocks are contemporaneously correlated.

Since the measurement equations (10)-(14) are a function of $\{\xi_t, \xi_{t-1}, ..., \xi_{t-47}\}$, the state equation should describe the joint distribution of ξ_t's from $t - 47$ to t, where I and 0 denote a (4×4) identity and zero matrix, respectively:

$$\underbrace{\begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \xi_{t-2} \\ \vdots \\ \xi_{t-46} \\ \xi_{t-47} \end{bmatrix}}_{192 \times 1} = \underbrace{\begin{bmatrix} \underbrace{I}_{4\times4} & \underbrace{0}_{4\times4} & 0 & 0 & ... & 0 & 0 & 0 \\ I & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & I & 0 & 0 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & ... & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & I & 0 & 0 \\ 0 & 0 & 0 & 0 & ... & 0 & I & 0 \end{bmatrix}}_{192 \times 192} \underbrace{\begin{bmatrix} \xi_{t-1} \\ \xi_{t-2} \\ \xi_{t-3} \\ \vdots \\ \xi_{t-47} \\ \xi_{t-48} \end{bmatrix}}_{192 \times 1} + \underbrace{\begin{bmatrix} \underbrace{\varepsilon_t}_{4\times1} \\ \underbrace{0}_{4\times1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_{192 \times 1}. \tag{17}$$

---

[21] The identical Kalman filter equations that emerge from an assumption of Normality can also be motivated using a least-squares criterion; see for example Hamilton (1994, Chapter 13).

2.2 Estimation

Our system takes the form of a nonlinear state space model in which the state transition equation is given by (17) and observation equation by (10)-(14) where $P_{it}(j)$ is given by (15) and $p_{it}(\tau)$ by (8). Our baseline model has 12 parameters to estimate, namely the diagonal terms in the variance matrices Σ and R and the parameters governing genuine duration dependence, δ_1, δ_2 and δ_3. Because the observation equation is nonlinear in x_{it}, the extended Kalman filter can be used to approximate the likelihood function for the observed data $\{y_1, ..., y_T\}$ and form an inference about the unobserved latent variables $\{\xi_1, ..., \xi_T\}$, as detailed in Appendix B. Inference about historical values for ξ_t provided below correspond to full-sample smoothed inferences, denoted $\hat{\xi}_{t|T}$.

3 Results for the baseline specification

We estimated parameters for the above nonlinear state-space model using seasonally adjusted monthly data on $y_t = (U_t^1, U_t^{2.3}, U_t^{4.6}, U_t^{7.12}, U_t^{13.+})'$ for $t =$ January 1976 through June 2017. Figure 4A plots smoothed estimates for $p_{it}(1)$, the probability that a newly unemployed worker of type i at $t-1$ will still be unemployed at t. These average 0.35 for type H individuals and 0.82 for type L individuals, close to the average calculations of 0.36 and 0.85, respectively, that we arrived at in row 2 of Table 1 when we were explaining the intuition behind our approach using steady-state calculations. The probabilities of type H individuals remaining unemployed rise during the early recessions but are less cyclical in the last two recessions. By contrast, the continuation probabilities for type L individuals rise in all recessions. The gap between the two probabilities increased significantly over the last 20 years.

Figure 4B plots inflows of individuals of each type into the pool of newly unemployed. Type H workers constitute 77% on average of the newly unemployed, again close to the value of 78% expected on the basis of the simple steady-state calculations in row 5 of Table 1. Inflows of both types increase during recessions. New inflows of type H workers declined immediately at the end of every recession, but inflows of type L workers continued to rise after the recessions of 1990-91 and 2001 and were still at above-average levels 3 years after the end of the Great Recession. This changing behavior of type L workers' inflows appears to be another important characteristic of jobless recoveries. The Great Recession is unique in that the inflows of type L workers as well as

14

the continuation probabilities reached higher levels than any earlier dates in our data set.

The combined implications of these cyclical patterns are summarized in Figure 5. Before the Great Recession, the share in total unemployment of type L workers fluctuated between 30% and 60%, falling during expansions and rising during and after recessions. But during the Great Recession, the share of type L workers skyrocketed to over 80%. The usual recovery pattern of a falling share of type L workers has been very slow in the aftermath of the Great Recession.

While the inflows of type H workers show a downward trend since the 1980's, those of type L workers exhibit an upward trend. This difference in the low frequency movements of the two series provides a new perspective on the secular decrease in the inflows to unemployment and the secular rise in the average duration of unemployment. Figure 4B shows that the downward trend in the inflows is mainly driven by type H workers. The increased share of type L inflows contributed to the rise in the average duration of unemployment since the 1980's. This suggests that unobserved heterogeneity is important in accounting for low frequency dynamics in the labor market as well as those for business cycle frequencies.

Table 2 provides parameter estimates for our baseline model. The estimated genuine duration dependence parameters, $\tilde{\delta}_1, \tilde{\delta}_2$, and $\tilde{\delta}_3$ are consistent with the scarring hypothesis– the longer someone from either group has been unemployed, provided the duration has been 11 months or less, the more likely it is that person will be unemployed next month. Once someone has been unemployed for more than a year, it becomes more likely as more months accumulate that they will either find a job or exit the labor force in any given month. This non-monotonic behavior of genuine duration dependence is displayed graphically in Panel A of Figure 6.

As seen in Panel B of Figure 6, our estimates of genuine duration dependence imply relatively modest changes in continuation probabilities for type L workers for most horizons. And while the implications for long-horizon continuation probabilities for type H workers may appear more significant, they are empirically irrelevant, since the probability that type H workers would be unemployed for more than 12 months is so remote. To gauge the overall significance of genuine duration dependence, we calculated the unemployment level predicted by our model for each date t in the sample if the values of $\tilde{\delta}_1, \tilde{\delta}_2$, and $\tilde{\delta}_3$ were all set to zero, and found it would only be about 4% lower on average than the value predicted by our baseline model. Thus although the values of $\tilde{\delta}_1$ and $\tilde{\delta}_3$ are statistically significant, they play a relatively minor role compared to ex ante heterogeneity

15

in accounting for differences in continuation probabilities by duration of unemployment.

3.1 Variance decomposition

Many previous studies have tried to summarize the importance of different factors in determining unemployment by looking at correlations between the observed unemployment rate and the steady-state unemployment rate predicted by each factor of interest alone; see for example Fujita and Ramey (2009) and Shimer (2012). One major benefit of our framework is that it delivers a much cleaner answer to this question in the form of variance decompositions, which measure how much each shock contributes to the mean squared error (MSE) of an s-period-ahead forecast of a magnitude of interest.[22]

Our model can be used to account for the difference between the unemployment realization at time $t + s$ and a forecast based on values of the state vector only through date t in terms of the sequence of shocks between t and $t + s$, denoted $\varepsilon_{t+1}, \varepsilon_{t+2}, ..., \varepsilon_{t+s}$. It is convenient to work with a linear approximation to that decomposition, which we show in Appendix C takes the form

$$y_{t+s} - \hat{y}_{t+s|t} \simeq \sum_{j=1}^{s} [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]\varepsilon_{t+j} \tag{18}$$

for $\Psi_{s,j}(\cdot)$ a known (5×4)-valued function of $\xi_t, \xi_{t-1}, ..., \xi_{t-47+j}$. The mean squared error matrix associated with an s-period-ahead forecast of y_{t+s} is then

$$
\begin{aligned}
E(y_{t+s} - \hat{y}_{t+s|t})(y_{t+s} - \hat{y}_{t+s|t})' &= \sum_{j=1}^{s} [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]\Sigma[\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]' \\
&= \sum_{j=1}^{s} \sum_{m=1}^{4} \Sigma_m [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})e_m][\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})e_m]'
\end{aligned}
\tag{19}
$$

for e_m column m of the (4×4) identity matrix and Σ_m the row m, column m element of Σ. Thus the contribution of innovations of type L worker's inflows (the first element of $\varepsilon_t = (\varepsilon_{Lt}^w, \varepsilon_{Ht}^w, \varepsilon_{Lt}^x, \varepsilon_{Ht}^x)'$) to the MSE of the s-period-ahead linear forecast error of total unemployment, $\iota_5' y_t$, is given by

---

[22] Note as in Den Haan (2000) and Hamilton (forthcoming) that an s-period-ahead forecast error can be stationary even if the unemployment rate is nonstationary, allowing us to calculate the contribution of different shocks to the MSE at any horizon s. This is another advantage of our approach over that seen in the earlier literature.

$$\iota_5' \sum_{j=1}^{s} \Sigma_1 [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j}) e_1][\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j}) e_1]' \iota_5 \tag{20}$$

where ι_5 denotes a (5×1) vector of ones. Note that as in the constant-parameter linear case, the sum of the contributions of the 4 different structural shocks would be equal to the MSE of an s-period-ahead linear forecast of unemployment in the absence of measurement error. However, in our case the linearization is taken around time-varying values of $\{\xi_t, \xi_{t-1}, ..., \xi_{t-47+j}\}$. We can evaluate equation (20) at the smoothed inferences $\{\hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t-47+j|T}\}$ and then take the average value across all dates t in the sample. This gives us an estimate of the contribution of the type L worker's inflows to unemployment fluctuations over a horizon of s months:

$$q_{s,1} = T^{-1} \sum_{t=1}^{T} \iota_5' \sum_{j=1}^{s} \Sigma_1 [\Psi_{s,j}(\hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t-47+j|T}) e_1][\Psi_{s,j}(\hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t-47+j|T}) e_1]' \iota_5. \tag{21}$$

Consequently $q_{s,1} / \sum_{m=1}^{4} q_{s,m}$ would be the ratio of the first factor's contribution to unemployment volatility at horizon s.

Figure 7 shows the contribution of each factor to the mean squared error in predicting overall unemployment as a function of the forecasting horizon. If one is trying to forecast unemployment one month ahead, uncertainty about future inflows of type H and type L workers are equally important. However, as one looks farther into the future, the single most important source of uncertainty becomes inflows of new type L workers, followed by uncertainty about their outflows. Much of the MSE associated with a 2-year-ahead forecast of unemployment comes from not knowing when the next recession will begin or the current recession will end. For this reason, the MSE associated with 2-year-ahead forecasts is closely related to what some researchers refer to as the "business cycle frequency" in a spectral decomposition.[23] We conclude that type L inflows are the most important factor in unemployment dynamics at the business-cycle frequency.

Panel B of Figure 7 breaks these contributions separately into inflows and outflows. Both inflows and outflows are important. Inflows account for about 60% of the MSE in predicting unemployment for any forecasting horizon. We conclude that inflows are the most important factor in the variability of unemployment at every frequency.

---

[23] See Hamilton (forthcoming) for discussion of why the error associated with a two-year-ahead forecast might be interpreted as the business cycle component of the series.

3.2 Historical decomposition

A separate question of interest is how much of the realized variation over some historical episode came from particular structural shocks. As in (18) our model implies an estimate of the contribution of shocks to a particular observed episode, namely

$$y_{t+s} - \hat{y}_{t+s|t} \simeq \sum_{j=1}^{s} [\Psi_{s,j}(\hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t-47+j|T})]\hat{\varepsilon}_{t+j|T} \tag{22}$$

where $\hat{\varepsilon}_{t+j|T} = \hat{\xi}_{t+j|T} - \hat{\xi}_{t+j-1|T}$. From this equation, we can estimate for example the contribution of $\varepsilon^w_{L,t+1}, \varepsilon^w_{L,t+2}, ..., \varepsilon^w_{L,t+s}$ (the shocks to w_L between $t+1$ and $t+s$) to the deviation of the level of unemployment at $t+s$ from the value predicted on the basis of initial conditions at t:

$$\iota_5' \sum_{j=1}^{s} [\Psi_{s,j}(\hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t-47+j|T})]e_1\hat{\varepsilon}'_{t+j|T}e_1. \tag{23}$$

Figure 8 shows the contribution of each component to the realized unemployment rate in the last five recessions. In each panel, the solid line (labeled U_{base}) gives the change in the unemployment rate relative to the value at the start of the episode that would have been predicted on the basis of initial conditions. Typically an increase in the inflow of type L workers (whose contribution to total unemployment is indicated by the starred red curves) is the most important reason that unemployment rises during a recession. A continuing increase of these inflows even after the recession was over was an important factor in the jobless recoveries from the 1990 and 2001 recessions.

During the first 8 months of the Great Recession, changes in inflows and outflows of type L individuals were of equal importance in accounting for rising unemployment. But our model concludes that new inflows of type L individuals were the most important factor contributing to rising unemployment after July of 2008.

4   Corroboration using other data sources and methods

In this section we review in evidence that has led some researchers to conclude that inflows are unimportant, and explain why we believe that conclusion is mistaken. The solid black line in Figure 9 plots the unemployment rate u_t during the Great Recession. In steady state this would

be related to the inflow rate n_t and outflow rate o_t by[24]

$$u_t \simeq \frac{n_t}{n_t + o_t}. \tag{24}$$

As in Shimer (2012) we can ask what the path of unemployment since 2007:Q4 would have been if o_t had stayed fixed at its 2007:Q4 value while n_t varied as actually observed. This is plotted as the dotted black line in Figure 9, and seems to suggest that new inflows into unemployment had little to do with the Great Recession. Hall and Schulhofer-Wohl (2017) reached a similar conclusion using an expression describing unemployment as a function of current outflows and past inflows.

The first concern we have with this kind of exercise is that an increase in inflows statistically predicts future changes in the outflow rate. Consider a bivariate VAR for $(\Delta o_t, \Delta n_t)$:

$$\Delta o_t = c_o + \phi_{oo,1}\Delta o_{t-1} + \cdots + \phi_{oo,8}\Delta o_{t-8} + \phi_{on,1}\Delta n_{t-1} + \cdots + \phi_{on,8}\Delta n_{t-8} + \varepsilon_{ot} \tag{25}$$

$$\Delta n_t = c_n + \phi_{no,1}\Delta o_{t-1} + \cdots + \phi_{no,8}\Delta o_{t-8} + \phi_{nn,1}\Delta n_{t-1} + \cdots + \phi_{nn,8}\Delta n_{t-8} + \varepsilon_{nt}. \tag{26}$$

The null hypothesis that $\phi_{on,1} = \cdots = \phi_{on,8} = 0$ based on OLS estimation of equation (25) for $t = 1969$:Q3 to 2016:Q4 is rejected with a p-value below 10^{-6}. In other words, a great deal of the observed variation in outflows could have been predicted on the basis of earlier values of inflows.

One way to characterize the role of inflows in accounting for changes in the unemployment rate in a way that is consistent with this predictability is to add u_t as a third variable to the above VAR and examine historical decompositions. We did this using a Cholesky decomposition of the covariance matrix with Δo_t ordered first to give as much of the benefit to Shimer's view as possible. This historical decomposition attributes more than half of the Great Recession surge in unemployment to inflows, as seen in the solid green line in Figure 9.

It is not just the level of new inflows but also their composition that predict future outflows. One indication of changes in composition comes from new claims for unemployment insurance (UI). Only individuals who are unemployed through no fault of their own can file for UI, and not everyone who is eligible files a claim. Those who do file claims may represent a subset of the newly unemployed who are more likely to have skills or attributes that are currently less in demand, such

---

[24]We followed Shimer's method to calculate updated quarterly values for n_t and o_t.

as individuals whose jobs were replaced by machines or whose skills are in declining industries–people who would be characterized by our approach as type L individuals. Let Q_t denote the number of individuals who file new claims in the last week of quarter t and U_t^1 the BLS estimate of the number of individuals who became newly unemployed in the last month of quarter t. We added Δq_t for $q_t = Q_t/U_t^1$ as another variable to (25). The hypothesis that coefficients on lags of Δq_{t-j} are all zero is again rejected ($p < 10^{-6}$). Changes in the composition of inflows have huge additional predictive power for future outflows beyond that contained in the level of inflows alone. If we add Δq_t as another variable to the VAR predicting unemployment, we would conclude that the level and composition of inflows account for most of the increase in unemployment during the Great Recession; see the dashed blue line in Figure 9.[25]

5 Robustness checks

Column 1 of Table 3 summarizes some of the key conclusions that emerge from our baseline analysis. The table breaks down the MSE of a forecast of the overall level of unemployment at 3-month, 1-year, and 2-year forecast horizons into the fraction of the forecast error that is attributable to various shocks. In our baseline model, inflows account for more than half the variance at all horizons. Inflows of type L workers are most important but the outflows of type L workers and the inflows of type H workers are also crucial at a 3-month horizon. At a 1- or 2-year horizon, shocks to inflows of type L workers are the single most important factor and shocks to outflow probabilities for type L workers are second most important factor. The table also reports asymptotic standard errors for each of these magnitudes.

Accounting for the structural break in the CPS. Our baseline estimates adjusted the unemployment duration data for the change in survey design in 1994 by using differences between rotation groups 1 and 5 and groups 2-4 and 6-8 in the CPS. Column 2 of Table 3 reports the analogous variance decompositions when we instead use Hornstein's (2012) data adjustment.[26] In column 3

---

[25] Several other recent studies have found that observable characteristics that were not considered in the earlier literature can further explain differences in unemployment durations. Faberman and Kudlyak (2017) discovered that how much time newly unemployed devote to searching for a new job predicts how long they will remain unemployed. Kudlyak and Lange (2017) demonstrated that a newly unemployed individual who the previous month had been classified as not in the labor force is likely to remain unemployed longer than a newly unemployed individual who had been employed the previous month. And Morchio (2016) documented that 2/3 of prime-age unemployment comes from only 10% of the workers.

[26] Note that although we report the log likelihood and Schwarz's (1978) Bayesian criterion in rows 2 and 3 of Table

we use only data subsequent to the redesign in 1994 making no adjustment to the reported BLS figures. Column 4 uses the full data set from 1976-2013 with no adjustments for the 1994 redesign (column 4). All specifications lead to the conclusion that changes for type L workers account for most of the cyclical fluctuations in unemployment.

Time-varying genuine duration dependence. Our baseline specification assumed that the parameters $\delta_1, \delta_2,$ and δ_3 characterizing genuine duration dependence in equations (5) and (8) do not change over time. Column 5 of Table 3 reports results for a more general specification

$$d_{\tau t} = \tilde{\delta}_{1t}((\tau - 1)/48) + \tilde{\delta}_{2t}[2((\tau - 1)/48)^2 - 1] + \tilde{\delta}_{3t}[4((\tau - 1)/48)^3 - 3((\tau - 1)/48)]$$

where $\tilde{\delta}_{jt} = \tilde{\delta}_j^{(1)}$ in normal months and $\tilde{\delta}_{jt} = \tilde{\delta}_j^{(2)}$ if the national unemployment rate is above 6.5%, times when the labor market is in slack and it is likely that many job losers automatically became eligible for extended UI benefits. Adding 3 new parameters $(\tilde{\delta}_1^{(2)}, \tilde{\delta}_2^{(2)}, \tilde{\delta}_3^{(2)})$ to the model results in an increase in the log likelihood of 46.2, but does not change any of our core conclusions.

Allowing for correlated shocks. Our baseline specification assumed that the shocks to w_{Lt}, w_{Ht}, p_{Lt} and p_{Ht} were mutually uncorrelated. We estimated a generalization of the model to allow for nonzero correlations deriving from a factor structure for the innovations, $\varepsilon_t = \lambda F_t + u_t$, where $F_t \sim N(0, 1)$, λ is a (4×1) vector of factor loadings, and u_t is a (4×1) vector of mutually uncorrelated idiosyncratic components with diagonal variance matrix $E(u_t u_t') = Q$, so $E(\varepsilon_t \varepsilon_t') = \lambda \lambda' + Q$. Note that equation (22) continues to hold in this more general setting, and we could still calculate the magnitude in (23), which measures what would happen if $\{\varepsilon_{L,t+j}^w\}_{j=1}^s$ were to have followed its inferred historical path with $\{\varepsilon_{H,t+j}^w, \varepsilon_{L,t+j}^x, \varepsilon_{H,t+j}^x\}_{j=1}^s$ all zero. This calculation would no longer have a clean statistical interpretation as the answer to a forecasting question when the ε's are correlated, because in the latter case knowledge of the value of one of the ε's would cause one to revise the contemporaneous forecast of the others. Nevertheless, we can still calculate the magnitude in (23) for the factor model as a check on whether the quantitative importance of type L inflows is in any way an artifact of having assumed uncorrelated shocks. The lower right panel of Figure 8 plots the quantitative contribution calculated in this way for each of the four shocks during the Great Recession. The graph is virtually identical to that in the lower left from our

---

3, the values for columns 2-4 are not comparable with the others due to a different definition of the observable data vector y_t.

baseline model.

We can also calculate the separate statistical contribution of each of the 5 uncorrelated shocks in the factor model, which consist of the aggregate factor F_t and the four elements of u_t. The contributions of each of the five shocks are summarized in column 6 of Table 3. The aggregate factor by itself accounts for 64% of the MSE of a 1–year-ahead forecast of unemployment. But the component of inflows of type L workers that is uncorrelated with the aggregate factor would still by itself account for 31% of the MSE, far more important than any other idiosyncratic shock. We conclude that the importance of inflows of type L workers is robust to assumptions about correlations between the shocks.

Time aggregation. Focusing on monthly transition probabilities misses people who lose their job in week 1 of a month but find a new job in week 2. We discuss some of the literature on this in Appendix E, and explain our reasons for favoring the specification in our baseline model. Column 7 of Table 3 reports that if we allow for weekly transitions, we would find a modestly smaller role for inflows than in our baseline model. This is to be expected, since by construction it imputes some people who gain new jobs only to lose them again before the month is over. Note however that the weekly model in column 7 has a slightly worse fit to the data than the baseline monthly model in column 1.

6   Conclusion

Representative worker models of unemployment can give rise to profoundly different dynamics from models that allow for heterogeneity. In this paper we demonstrated how to estimate a dynamic model of unemployment allowing for unobserved heterogeneity, and concluded that new inflows of individuals with low job-finding probabilities are the dominant feature of economic recessions.

22

References

Aaronson, Daniel, Bhashkar Mazumder, and Shani Schechter (2010). "What is behind the rise in long-term unemployment," *Economic Perspectives*, QII:28-51.

Ahn, Hie Joo (2016), "The Role of Observed and Unobserved Heterogeneity in the Duration of Unemployment Spells," Working paper, Federal Reserve Board.

Alvarez, Fernando, Katarína Borovičková, and Robert Shimer (2016). "Decomposing Duration Dependence in a Stopping Time Model," NBER Working Papers 22188, National Bureau of Economic Research, Inc.

Bachmann, Ronald, and Mathias Sinning (2016). "Decomposing the Ins and Outs of Cyclical Unemployment," *Oxford Bulletin of Economics and Statistics,* 78(6): 853-876.

Baker, Michael (1992). "Unemployment Duration: Compositional Effects and Cyclical Variability," *American Economic Review*, 82(1):313–21.

Barnichon, Regis and Andrew Figura (2015). "Labor Market Heterogeneity and the Aggregate Matching Function," *American Economic Journal: Macroeconomics,* 7(4): 222-49.

Baumeister, Christiane and Gert Peersman (2013). "Time-Varying Effects of Oil Supply Shocks on the U.S. Economy," *American Economic Journal: Macroeconomics*, 5(4): 1-29.

Den Haan, Wouter J. (2000). "The Comovement between Output and Prices," *Journal of Monetary Economics*, 46:3-30.

Elsby, Michael W. L., Ryan Michaels, and Gary Solon (2009). "The Ins and Outs of Cyclical Unemployment," *American Economic Journal: Macroeconomics*, 1(1): 84-110.

Faberman, R. Jason and Marianna Kudlyak (2017). "The Intensity of Job Search and Search Duration," Working Paper, Federal Reserve Bank of Chicago.

Fujita, Shigeru and Garey Ramey (2009). "The Cyclicality of Separation and Job Finding Rates," *International Economic Review*, 50(2):415-430.

Hall, Robert E. (2005). "Job Loss, Job Finding and Unemployment in the U.S. Economy over the Past Fifty Years," in Mark Gertler and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2005*, Volume 20, National Bureau of Economic Research, Inc.

Hall, Robert E. (2014). "Quantifying the Lasting Harm to the U.S. Economy from the Financial Crisis," in Jonathan Parker and Michael Woodford, eds., *NBER Macroeconomics Annual* 2014.

Volume 29, National Bureau of Economic Research, Inc.

Hall, Robert E. and Sam Schulhofer-Wohl (2017). "The Pervasive Importance of Tightness in Labor-Market Volatility," working paper, Stanford University.

Ham, John C., and Samuel A. Rea, Jr. (1987). "Unemployment Insurance and Male Unemployment Duration in Canada," *Journal of Labor Economics*, 5(3): 325-353.

Hamilton, James D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, James D. (forthcoming). "Why You Should Never Use the Hodrick-Prescott Filter," *Review of Economics and Statistics*.

Heckman, James and B. Singer (1984a). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2): 271-320.

Heckman, James, and B Singer (1984b), "Econometric duration analysis", *Journal of Econometrics*, 24: 63-132.

Honoré, Bo E. (1993). "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60(1): 241-246.

Hornstein, Andreas (2012). "Accounting for Unemployment: the Long and Short of it," FRB Richmond Working Paper, No.12-07, Federal Reserve Bank of Richmond.

Ilg, Randy E., and Eleni Theodossiou (2012). "Job Search of the Unemployed by Duration of Unemployment," *Monthly Labor Review,* March 2012, pp.41-49, Bureau of Labor Statistics.

Katz, Lawrence F. and Bruce D. Meyer (1990). "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes," *The Quarterly Journal of Economics,* 105(4):973-1002.

Kudlyak, Marianna, and Fabian Lange (2014). "Measuring Heterogeneity in Job Finding Rates Among the Nonemployed Using Labor Force Status Histories," working paper, Federal Reserve Bank of Richmond.

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

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, and Lawrence F. Katz (2016). "Long-Term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Non-Participation," *Journal of Labor Economics*, 34:S7-S54.

Morchio, Iacopo (2016). "Work Histories and Lifetime Unemployment," working paper, University of Vienna.

Schwarz, Gideon (1978). "Estimating the Dimension of a Model," *Annals of Statistics* 6:461-464.

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

Van den Berg, Gerald J. (2001). "Duration Models: Specification, Identification, and Multiple Durations," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Volume V (North-Holland, Amsterdam).

Van den Berg, Gerald J., and Bas van der Klaauw (2001). "Combining Micro and Macro Unemployment Duration Data," *Journal of Econometrics* 102: 271-309.

Van den Berg, Gerald J., and Jan C. van Ours (1996). "Unemployment Dynamics and Duration Dependence," *Journal of Labor Economics*, 14(1):100-125.

White, Halbert (1982). "Maximum Likelihood Estimation of Misspecified Models," *Econometrica,* 50:1-25.

Yashiv, Eran (2007). "US Labor Market Dynamics Revisited," *The Scandinavian Journal of Economics,* 109(4): 779–806.

Table 1. Actual and predicted values for unemployment on average and during Great Recession using different steady-state representations

	Parameter values					Actual or predicted values				
						U^1	$U^{2.3}$	$U^{4.6}$	$U^{7.12}$	$U^{13.+}$
	1976:M1-2017:M06					**Observed values**				
(1)						**3,178**	**2,281**	**1,244**	**1,064**	**664**
	w	p				Fitted (and predicted) values				
(2)	3,178	0.4840				3,178	2,281	(618)	(78)	(1)
	w_H	w_L	p_H	p_L		Fitted (and predicted) values				
(3)	2,488	690	0.3559	0.8476		3,178	2,281	1,244	1,064	(621)
	w	δ_0	δ_1	δ_2	δ_3	Fitted values				
(4)	3,178	0.1090	-0.3690	0.0140	3.314e-5	3,178	2,281	1,244	1,064	664
	w_H	w_L	$p_H(1)$	$p_L(1)$	δ	Fitted values				
(5)	2,482	696	0.3550	0.8440	-0.0050	3,178	2,281	1,244	1,064	664
	2007:M12-2013:M12					**Observed values**				
(6)						**3,339**	**2,787**	**2,131**	**2,426**	**1,902**
	w_H	w_L	p_H	p_L		Fitted (and predicted) values				
(7)	2,274	1,065	0.32920	0.890		3,339	2,787	2,131	2,426	(2,358)
	w	δ_0	δ_1	δ_2	δ_3	Fitted values				
(8)	3,339	0.2360	-0.6620	0.0540	-1.27e-3	3,339	2,787	2,131	2,426	1,902
	w_H	w_L	$p_H(1)$	$p_L(1)$	δ	Fitted values				
(9)	2,307	1,032	0.3340	0.9000	0.0170	3,339	2,787	2,131	2,426	1,902

Notes to Table 1. Table reports average values of U_t^x in thousands of workers over the entire 1976:M1-2017:M6 sample and the 2007:M12-2013:M12 subsample along with predicted values from simple steady-state calculations. Parameters were chosen to fit exactly the values in that row appearing in normal face, while the model's predictions for other numbers are indicated by parentheses.

Table 2. Parameter estimates for the baseline model

σ_L^w	0.0422***	R_1	0.1011***	$\tilde{\delta}_1$	5.0512***
	(0.0039)		(0.0054)		(1.9164)
σ_H^w	0.0437***	$R_{2.3}$	0.0753***	$\tilde{\delta}_2$	-0.0485
	(0.0057)		(0.0044)		(0.0532)
σ_L^x	0.0476***	$R_{4.6}$	0.0817***	$\tilde{\delta}_3$	2.1674***
	(0.0054)		(0.0073)		(0.8104)
σ_H^x	0.0204***	$R_{7.12}$	0.0586***		
	(0.0027)		(0.0047)		
		R_{13+}	0.0393***		
			(0.0025)		
No. of Obs.		498			
Log-Likelihood		2,574.0			

Notes to Table 2. White (1982) quasi-maximum-likelihood standard errors in parentheses. See footnote 20 for the definition of $\tilde{\delta}_j$.

Table 3. Comparison of variance decomposition across different models

	Shocks	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
# of Param.		12	12	12	12	15	16	12	
Log-L.		2574.0	2474.7	1326.6	2608.7	2620.2	2585.7	2571.0	
SIC		-5073.4	-4874.8	-2585.5	-5142.8	-5147.2	-5072.0	-5067.5	
3 month	F	-	-	-	-	-	0.588	-	
		-	-	-	-	-	(0.069)	-	
	w_L	0.415	0.414	0.225	0.131	0.392	0.233	0.384	
		(0.045)	(0.048)	(0.053)	(0.025)	(0.046)	(0.041)	(0.044)	
	w_H	0.208	0.229	0.225	0.388	0.214	0.123	0.147	
		(0.041)	(0.048)	(0.057)	(0.052)	(0.044)	(0.030)	(0.040)	
	p_L	0.281	0.288	0.279	0.203	0.281	0.000	0.253	
		(0.043)	(0.046)	(0.067)	(0.039)	(0.045)	(0.055)	(0.041)	
	p_H	0.096	0.068	0.271	0.277	0.114	0.057	0.216	
		(0.022)	(0.017)	(0.067)	(0.045)	(0.022)	(0.016)	(0.046)	
	Inflows	0.623	0.644	0.450	0.519	0.606	0.355	0.531	
	L group	0.696	0.703	0.503	0.335	0.673	0.233	0.637	
1 year	F	-	-	-	-	-	0.635	-	
		-	-	-	-	-	(0.070)	-	
	w_L	0.510	0.489	0.350	0.307	0.489	0.297	0.476	
		(0.050)	(0.052)	(0.060)	(0.042)	(0.052)	(0.051)	(0.049)	
	w_H	0.074	0.083	0.102	0.210	0.079	0.045	0.051	
		(0.017)	(0.022)	(0.030)	(0.041)	(0.020)	(0.012)	(0.016)	
	p_L	0.380	0.403	0.399	0.318	0.388	0.000	0.391	
		(0.051)	(0.053)	(0.071)	(0.053)	(0.054)	(0.071)	(0.053)	
	p_H	0.036	0.025	0.148	0.165	0.044	0.022	0.082	
		(0.009)	(0.007)	(0.045)	(0.034)	(0.010)	(0.006)	(0.022)	
	Inflows	0.584	0.572	0.453	0.516	0.567	0.343	0.527	
	L group	0.890	0.892	0.750	0.625	0.877	0.298	0.867	
2 year	F	-	-	-	-	-	0.644	-	
		-	-	-	-	-	(0.072)	-	
	w_L	0.523	0.485	0.427	0.430	0.503	0.310	0.483	
		(0.052)	(0.053)	(0.064)	(0.048)	(0.054)	(0.053)	(0.051)	
	w_H	0.049	0.054	0.063	0.139	0.053	0.031	0.033	
		(0.012)	(0.015)	(0.020)	(0.031)	(0.014)	(0.008)	(0.012)	
	p_L	0.403	0.444	0.413	0.320	0.413	0.000	0.428	
		(0.053)	(0.055)	(0.071)	(0.050)	(0.056)	(0.075)	(0.054)	
	p_H	0.025	0.017	0.096	0.112	0.031	0.015	0.055	
		(0.006)	(0.005)	(0.032)	(0.026)	(0.007)	(0.004)	(0.017)	
	Inflows	0.572	0.539	0.491	0.568	0.556	0.341	0.516	
	L group	0.926	0.929	0.841	0.750	0.916	0.310	0.911	

Notes to Table 3. (1) Baseline model, (2) alternative data, (3) post 94 data, (4) unadjusted data, (5) time-varying GDD, (6) correlated shocks, (7) weekly frequency. Standard errors in parentheses.
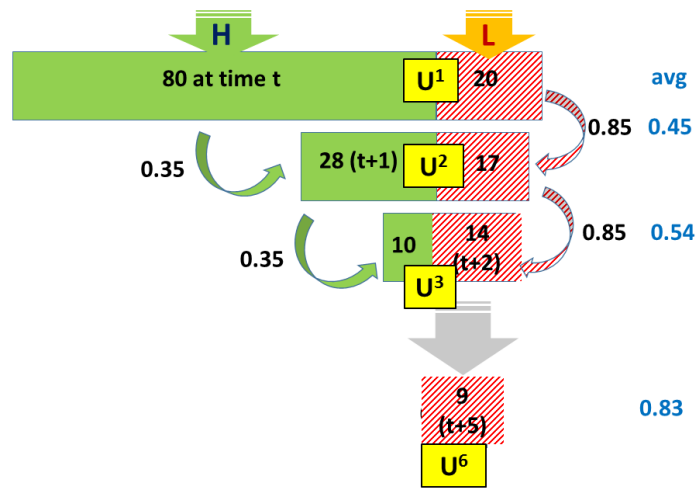
Figure 1. Illustration of how ex ante heterogeneity can cause unemployment-continuation probabilities to increase with duration.

Notes to Figure 1. Of newly unemployed at time t, 80 have unemployment-continuation probability of 35% and 20 have probability of 85%. The figure reports the number from each group who are still unemployed in subsequent months and the average continuation probabilities for each surviving cohort.
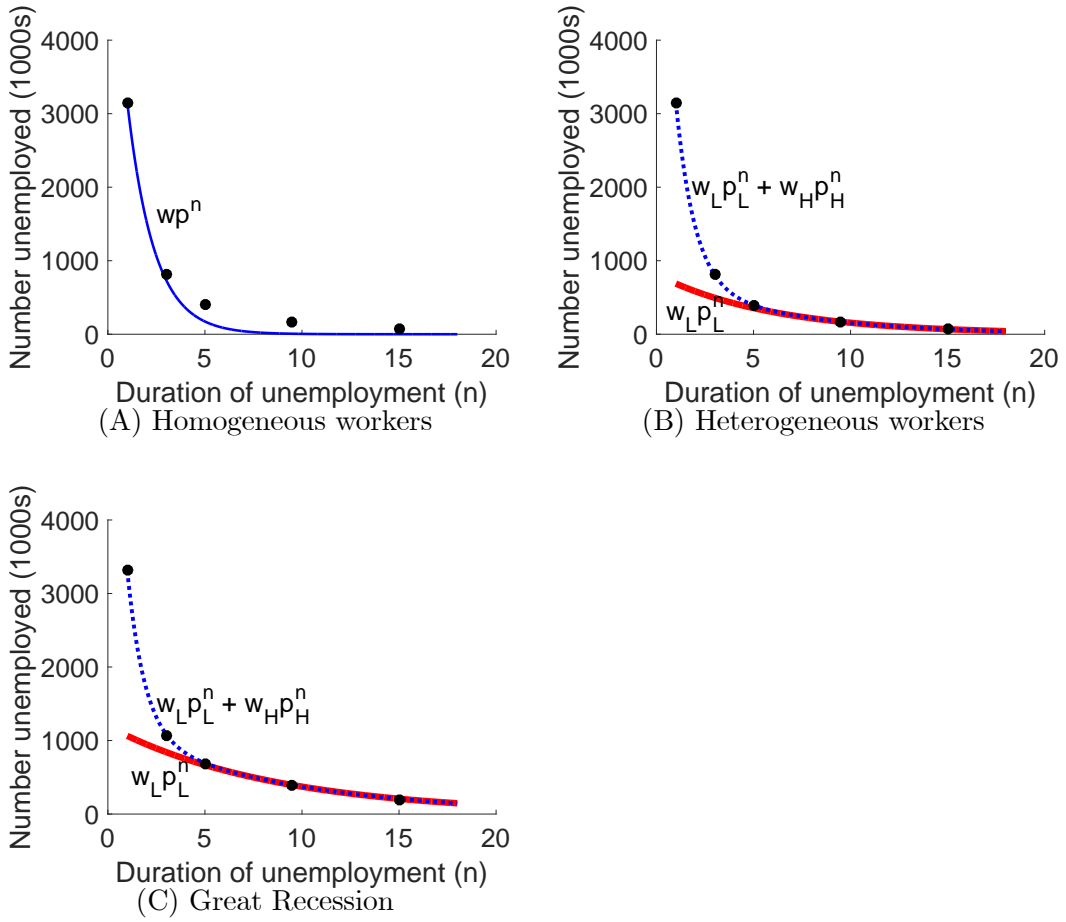
Figure 2. Predicted (smooth curves) and actual (black circles) numbers of unemployed as a function of duration based on constant-parameter specifications

Notes to Figure 2. Horizontal axis shows duration of unemployment in months and vertical axis shows number of unemployed for that duration in thousands of individuals. Circles denote imputed values for $\bar{U}^1, \bar{U}^3, \bar{U}^5$, $\bar{U}^{9.5}$, and $\bar{U}^{15}$ based on equation (6) with w_L, w_H, x_L, x_H, and δ chosen to fit the observed values of $\bar{U}^1$, $\bar{U}^{2.3}$, $\bar{U}^{4.6}$, $\bar{U}^{7.12}$, and $\bar{U}^{13.+}$ exactly. Panel A: homogeneous specification fit to 1976:M1-2017:M6 historical averages for $\bar{U}^1$, and $\bar{U}^{2.3}$. Panel B: pure cross-sectional heterogeneity specification fit to 1976:M1-2017:M6 historical averages for $\bar{U}^1$, $\bar{U}^{2.3}$, $\bar{U}^{4.6}$, and $\bar{U}^{7.12}$. Panel C: pure cross-sectional heterogeneity specification fit to average values for 2007:M12-2013:M12 for $\bar{U}^1$, $\bar{U}^{2.3}$, $\bar{U}^{4.6}$, and $\bar{U}^{7.12}$.

30

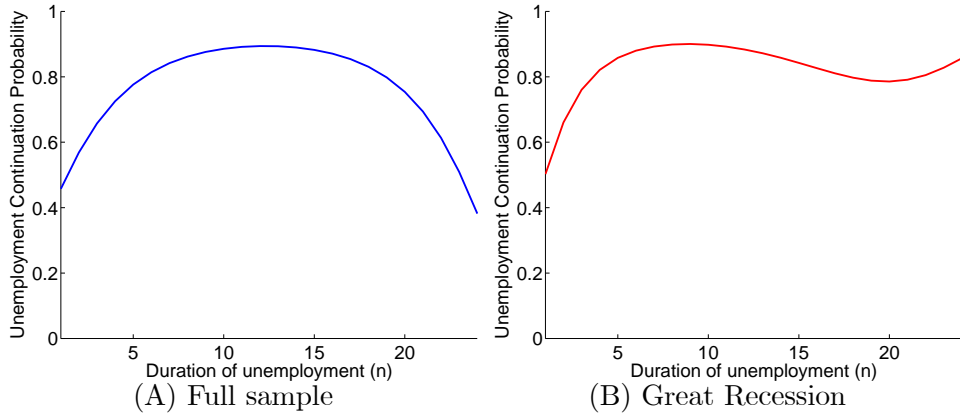(A) Full sample   (B) Great Recession

Figure 3. Unemployment-continuation probabilities as a function of duration based on constant-parameter pure genuine duration dependence specification

Notes to Figure 3. Horizontal axis shows duration of unemployment in months; vertical axis shows probability that individual is still unemployed the following month. Curves denote predicted values from the 5-parameter pure GDD model (w plus parameters in equation (5)) fit to 1976:M1-2017:M6 historical average values for $\bar{U}^1$, $\bar{U}^{2.3}$, $\bar{U}^{4.6}$, $\bar{U}^{7.12}$ and $\bar{U}^{13.+}$ (panel A) and for values for 2007:M12-2013:M12 (panel B). Each GDD model exactly fits the dots in Figure 2.

(A) Unemployment continuation probabilities   (B) Number of newly unemployed
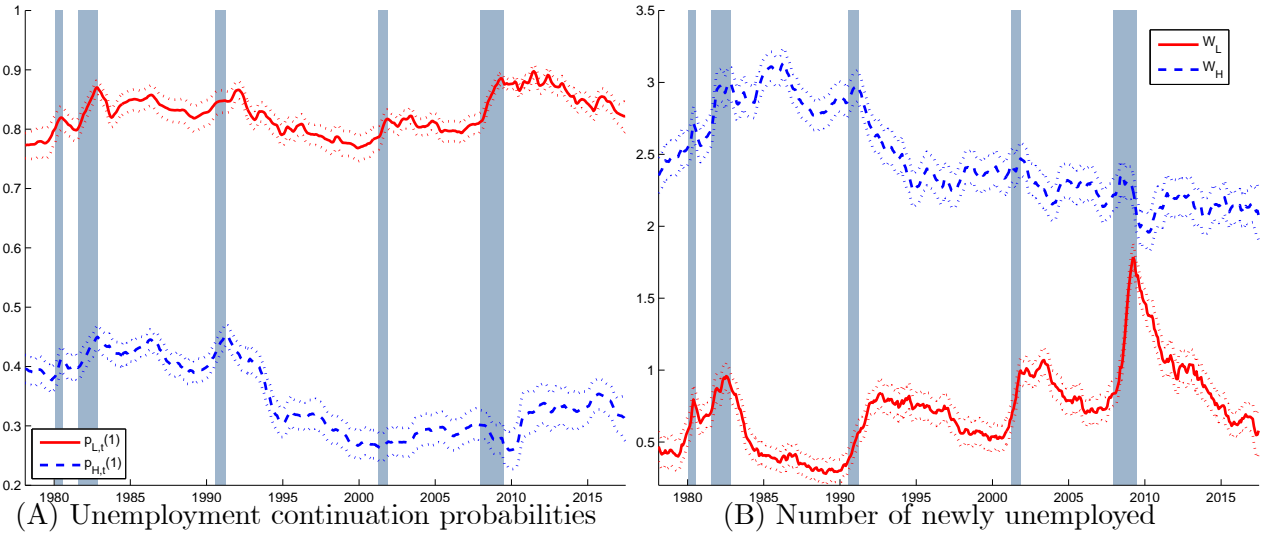
Figure 4. Probability that a newly unemployed worker will still be unemployed the following month and number of newly unemployed

Notes to Figure 4. Panel A plots $\hat{p}_{it|T}(1)$ for $i = L, H$ with 95% confidence intervals. Panel B plots $\hat{w}_{it|T}$ for $i = L, H$ with 95% confidence intervals.
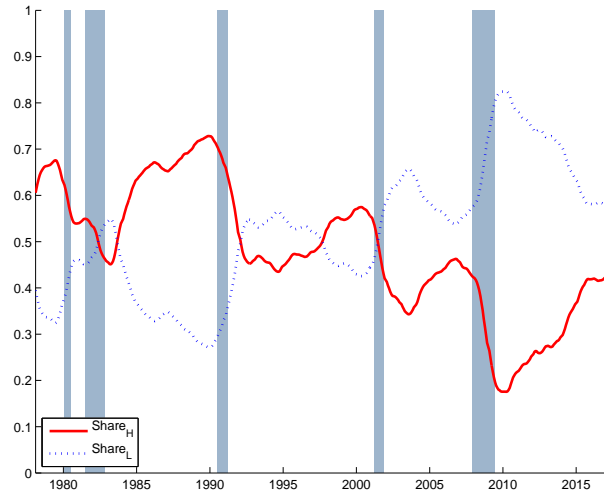
31

Figure 5. Share of total unemployment accounted for by each type of worker

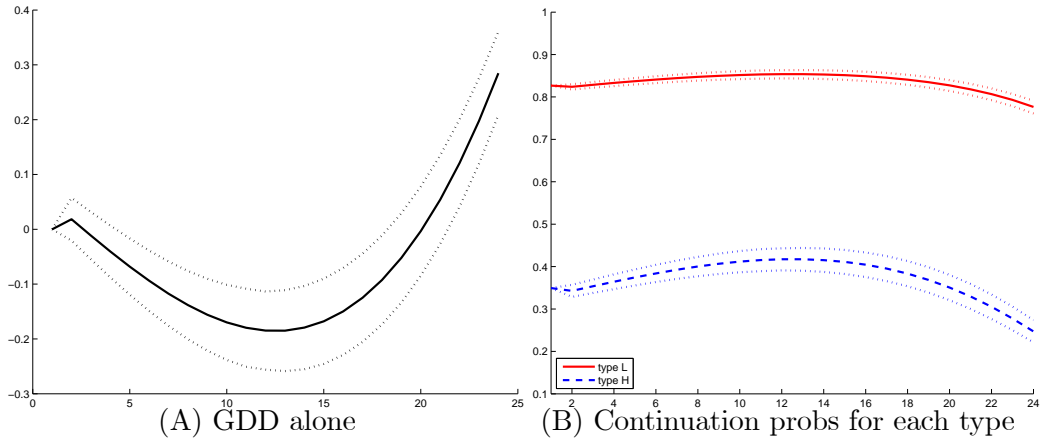(A) GDD alone   (B) Continuation probs for each type

Figure 6. Effects of genuine duration dependence

Notes to Figure 6. Panel A plots d_τ as a function of τ (months spent in unemployment). Panel B plots average unemployment-continuation probabilities of type H and type L workers as a function of duration of unemployment.
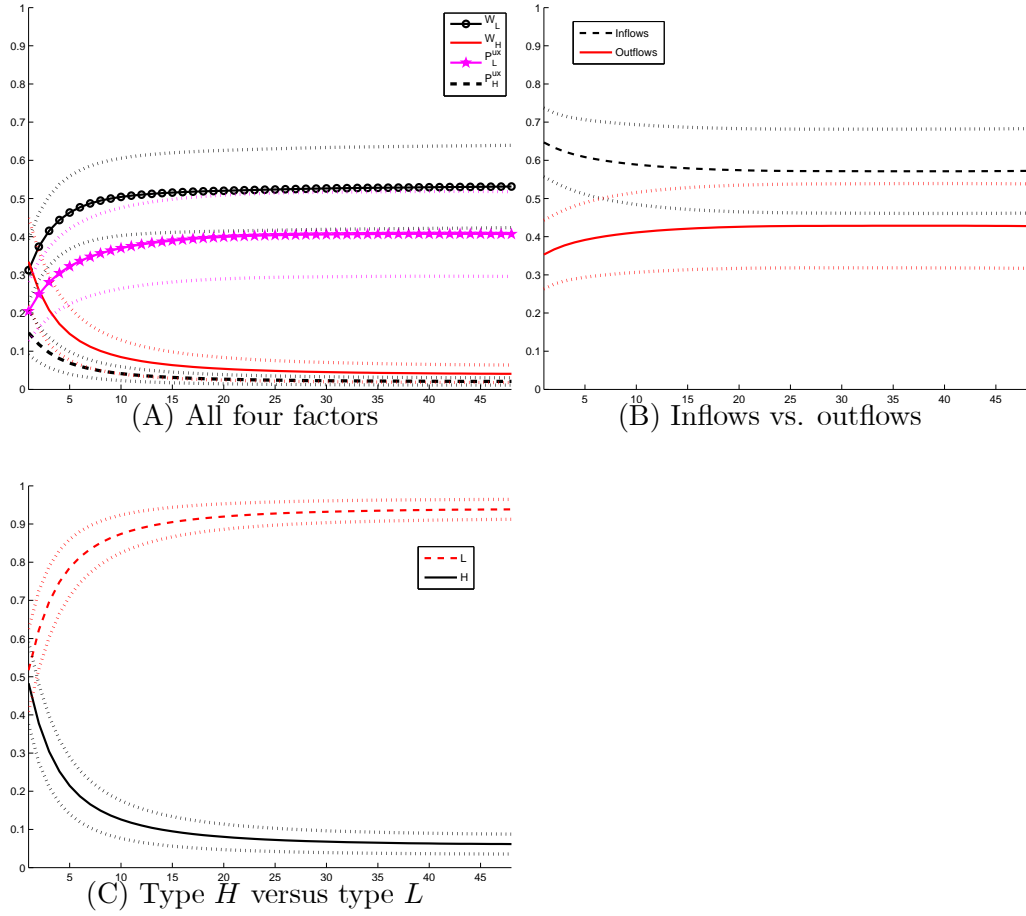
32

Figure 7. Fraction of variance of error in forecasting total unemployment at different horizons attributable to separate factors

Notes to Figure 7. Horizontal axis indicates the number of months ahead s for which the forecast is formed. Panel A plots the contribution of each of the factors $\{w_{Ht}, w_{Lt}, x_{Ht}, x_{Lt}\}$ separately, Panel B shows combined contributions of $\{w_{Ht}, w_{Lt}\}$ and $\{x_{Ht}, x_{Lt}\}$, and Panel C shows combined contributions of $\{w_{Ht}, x_{Ht}\}$ and $\{w_{Lt}, x_{Lt}\}$. Dotted lines denote 95% confidence intervals.
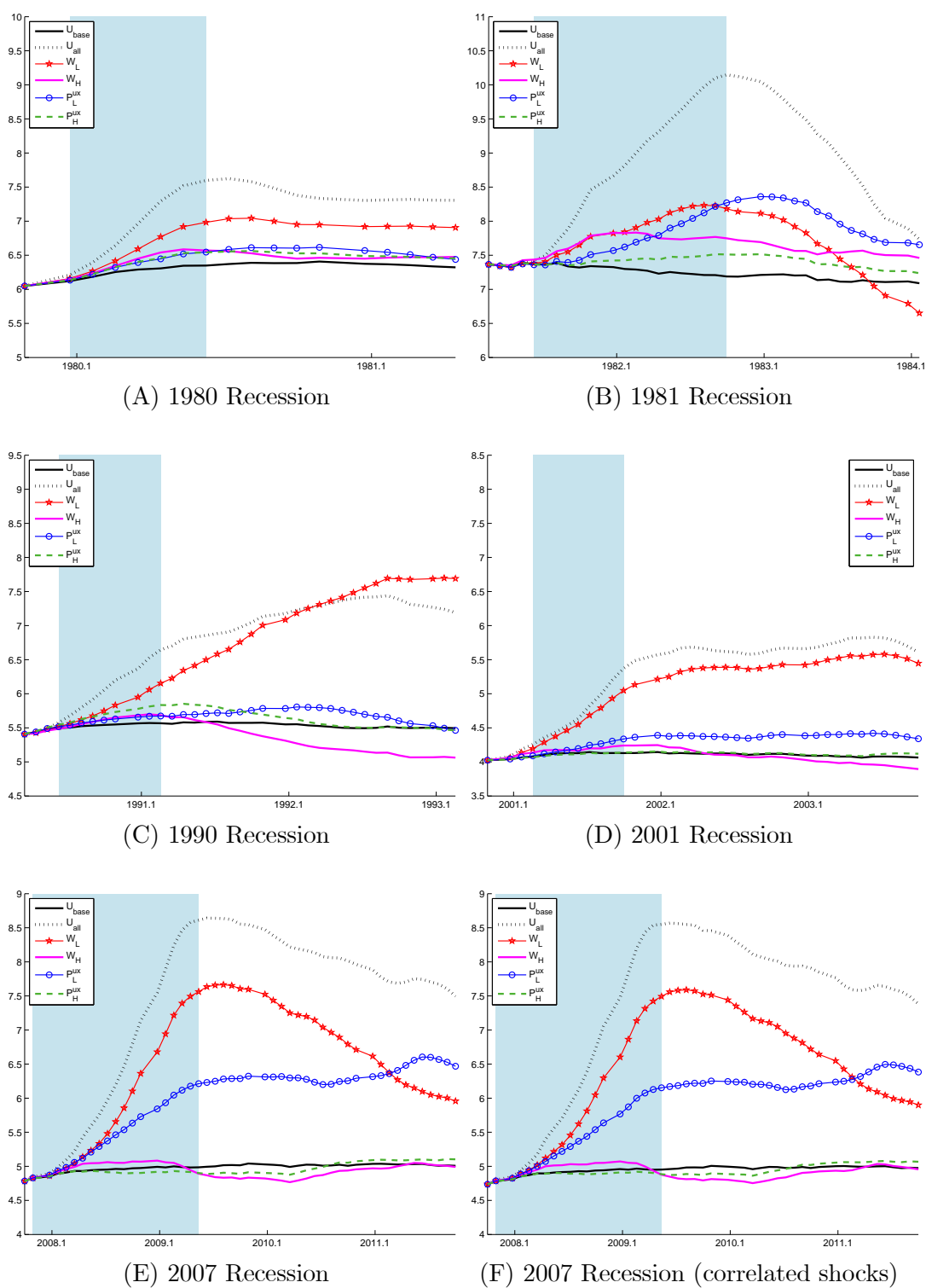
Figure 8. Historical decompositions of five U.S. recessions

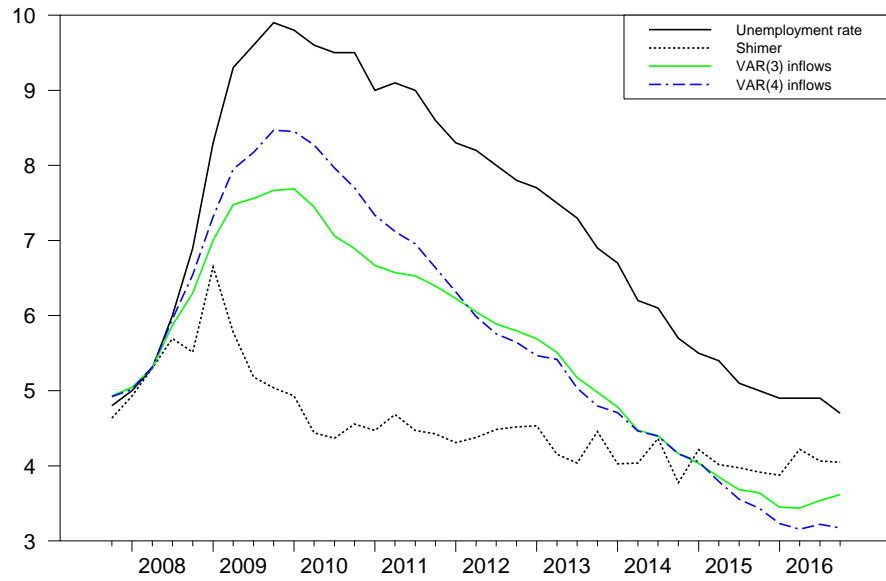Notes to Figure 8. Panels A-E use baseline model. Panel F uses factor model.

Figure 9. Contribution of inflows to level of unemployment

Notes to Figure 9. Solid black line: unemployment rate, 2007:Q4 to 2016:Q4. Dotted black line: -1.5 plus 100 times the value of equation (24) with o_t fixed at its 2007:Q4 value. Solid green line: value predicted by 3-variable VAR ordered o_t, n_t, u_t with shocks to o_t and u_t set to zero. Dashed blue line: value predicted by 4-variable VAR ordered o_t, n_t, q_t, u_t with shocks to o_t and u_t set to zero.

Online Appendix

A. Measurement issues and seasonal adjustment

The seasonally adjusted numbers of people unemployed for less than 5 weeks, for between 5 and 14 weeks, 15-26 weeks and for longer than 26 weeks are published by the Bureau of Labor Statistics. To further break down the number unemployed for longer than 26 weeks into those with duration between 27 and 52 weeks and with longer than 52 weeks, we used seasonally unadjusted CPS microdata publicly available at the NBER website (http://www.nber.org/data/cps_basic.html). Since the CPS is a probability sample, each individual is assigned a unique weight that is used to produce the aggregate data. From the CPS microdata, we obtain the number of unemployed whose duration of unemployment is between 27 and 52 weeks and the number longer than 52 weeks. We seasonally adjust the two series using X-12-ARIMA,[27] and calculated the ratio of those unemployed 27-52 weeks to the sum. We then multiplied this ratio by the published BLS seasonally adjusted number for individuals who had been unemployed for longer than 26 weeks to obtain our series $U_t^{7.12}$.[28]

An important issue in using these data is the redesign of the CPS in 1994. Before 1994, individuals were always asked how long they had been unemployed. After the redesign, if an individual is reported as unemployed during two consecutive months, then her duration is recorded automatically as the sum of her duration last month and the number of weeks between the two months' survey reference periods. Note that if an individual was unemployed during each of the two weeks surveyed, but worked at a job in between, that individual would likely self-report a duration of unemployment to be less than 5 weeks before the redesign, but the duration would be imputed to be a number greater than 5 weeks after the redesign.

As suggested by Elsby, Michaels and Solon (2009) and Shimer (2012) we can get an idea of the size of this effect by making use of the staggered CPS sample design. A given address is sampled for 4 months (called the first through fourth rotations, respectively), not sampled for the next 8

---

[27] An earlier version of this paper dealt with seasonality by taking 12-month moving averages and arrived at similar overall results to those presented in this version. As a further check on the approach used here, we compared the published BLS seasonally adjusted number for those unemployed with duration between 15 and 26 weeks to an X-12-ARIMA-adjusted estimate constructed from the CPS microdata, and found the series to be quite close.

[28] This adjustment is necessary because the published number for unemployed with duration longer than 26 weeks is different from that directly computed from the CPS microdata, although the difference is subtle. The difference arises because the BLS imputes the numbers unemployed with different durations to various factors, e.g., correction of missing observations.

months, and then sampled again for another 4 months (the fifth through eighth rotations). After the 1994 redesign, the durations for unemployed individuals in rotations 2-4 and 6-8 are imputed, whereas those in rotations 1 and 5 are self-reported, just as they were before 1994. For those in rotation groups 1 and 5, we can calculate the fraction of individuals who are newly unemployed and compare this with the total fraction of newly unemployed individuals across all rotations. The ratio of these two numbers is reported in Panel A of Figure A1, and averaged 1.15 over the period 1994-2007 as reported in the second row of Table A1. For comparison, the ratio averaged 1.01 over the period 1989-1993, as seen in the first row. This calculation suggests that if we want to compare the value of U_t^1 as calculated under the redesign to the self-reported numbers available before 1994, we should multiply the former by 1.15. This is similar to the adjustment factors of 1.10 used by Hornstein (2012), 1.154 by Elsby, Michaels and Solon (2009), 1.106 by Shimer (2012), and 1.205 by Polivka and Miller (1998).

For our study, unlike most previous researchers, we also need to specify which categories the underreported newly unemployed are coming from. Figure A1 reports the observed ratios of rotation 1 and 5 shares to the total for the various duration groups, with average values summarized in Table A1. One interesting feature is that under the redesign, the fraction of those with 7-12 month duration from rotations 1 and 5 is very similar to that for other rotations, whereas the fraction of those with 13 or more months is much lower.[29] Based on the values in Table A1, we should scale up the estimated values for U_t^1 and scale down the estimated values of $U_t^{2.3}$ and $U_t^{13.+}$ relative to the pre-1994 numbers. The values for $U_t^{4.6}$ and $U_t^{7.12}$ seem not to have been affected much by the redesign. Our preferred adjustment for data subsequent to the 1994 redesign is to multiply U_t^1 by 1.15, $U_t^{2.3}$ by 0.87, $U_t^{13.+}$ by 0.77, and leave $U_t^{4.6}$ and $U_t^{7.12}$ as is. We then multiplied all of our adjusted duration figures by the ratio of total BLS reported unemployment to the sum of our adjusted series in order to match the BLS aggregate exactly.

Hornstein (2012) adopted an alternative adjustment, assuming that all of the imputed newly unemployed came from the $U^{2.3}$ category. He chose to multiply U_t^1 by 1.10 and subtract the added workers solely from the $U_t^{2.3}$ category. As a robustness check we also report results using

---

[29]One possible explanation is digit preference– an individual is much more likely to report having been unemployed for 12 months than 13 or 14 months. When someone in rotation 5 reports they have been unemployed for 12 months, BLS simply counts them as such, and if they are still unemployed the following month, BLS imputes to them a duration of 13 months. The imputed number of people 13 months and higher is significantly bigger than the self-reported numbers, just as the imputed number of people with 2-3 months appears to be higher than self-reported.

Hornstein's proposed adjustment in Section 5.1, as well as results using no adjustments at all.

An alternative might be to to use the ratios for each t in Figure A1 rather than to use the averages from Table A1. However, as Shimer (2012) and Elsby, Michaels and Solon (2009) mentioned, such an adjustment would be based on only about one quarter of the sample and thus multiplies the sampling variance of the estimate by about four, which implies that noise from the correction procedure could be misleading in understanding the unemployment dynamics.

Table A1. Average ratio of each duration group's share in the first/fifth rotation group to that in total unemployment

	U^1	$U^{2.3}$	$U^{4.6}$	$U^{7.12}$	$U^{13.+}$
1989-1993	1.01	1.01	0.96	1.02	0.97
1994-2007	1.15	0.87	0.95	1.05	0.77

B. Estimation algorithm

The system in Section 2.1 can be written as

$$x_t = F x_{t-1} + v_t$$

$$y_t = h(x_t) + r_t$$

for $x_t = (\xi_t', \xi_{t-1}', ..., \xi_{t-47}')'$, $E(v_t v_t') = Q$, and $E(r_t r_t') = R$. The function $h(.)$ as well as elements of the variance matrices R and Q depend on the parameter vector $\theta = (\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3, R_1, R_{2.3}, R_{4.6}, R_{7.12}, R_{13+}, \sigma_L^w, \sigma_H^w, \sigma_L^x, \sigma_H^x)'$. The extended Kalman filter (e.g., Hamilton, 1994b) can be viewed as an iterative algorithm to calculate a forecast $\hat{x}_{t|t-1}$ of the state vector conditioned on knowledge of θ and observation of $Y_{t-1} = (y_{t-1}', y_{t-2}', ..., y_1')'$ with $P_{t|t-1}$ the MSE of this forecast. With these we can approximate the distribution of y_t conditioned on Y_{t-1} as $N(h(\hat{x}_{t|t-1}), \Omega_t)$ for $\Omega_t = H_t' P_{t|t-1} H_t + R$ and $H_t = \partial h(x_t)/\partial x_t'|_{x_t = \hat{x}_{t|t-1}}$ from which the approximate likelihood function associated with that θ,

$$\ell(\theta) = \sum_{t=1}^T \ln f(y_t | Y_{t-1}; \theta)$$

$$\ln f(y_t | Y_{t-1}; \theta) = -(1/2)\ln(2\pi) - (1/2)\ln|\Omega_t| - (1/2)[y_t - h(\hat{x}_{t|t-1})]'\Omega_t^{-1}[y_t - h(\hat{x}_{t|t-1})],$$

can be maximized numerically. The forecast of the state vector can be updated by iterating on

$$K_t = P_{t|t-1}H_t(H_t'P_{t|t-1}H_t + R)^{-1}$$

$$P_{t|t} = P_{t|t-1} - K_tH_t'P_{t|t-1}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - h(\hat{x}_{t|t-1}))$$

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t}$$

$$P_{t+1|t} = FP_{t|t}F' + Q.$$

Prior to the starting date January 1976 for our sample, BLS aggregates are available but not the micro data that we used to construct $U_t^{13.+}$. For the initial value for the extended Kalman filter, we calculated the values that would be implied if pre-sample values had been realizations from an initial steady state, estimating the (4×1) vector $\bar{\xi}_0$ from the average values for $\bar{U}^1, \bar{U}^{2.3}, \bar{U}^{4.6}$, and $\bar{U}^{7.+}$ over February 1972 - January 1976 using the method described in Section 1.1. Our initial guess was then $\hat{x}_{1|0} = \iota_{48} \otimes \bar{\xi}_0$ where ι_{48} denotes a (48×1) vector of ones. Diagonal elements of $P_{1|0}$ determine how much the presample values of ξ_j are allowed to differ from this initial guess $\hat{\xi}_{j|0}$. For this we set $E(\xi_j - \hat{\xi}_{j|0})(\xi_j - \hat{\xi}_{j|0})' = c_0I_4 + (1-j)c_1I_4$ with $c_0 = 10$ and $c_1 = 0.1$. The value for c_0 is quite large relative to the range of $\xi_{t|T}$ over the complete observed sample, ensuring that the particular value we specified for $\hat{x}_{1|0}$ has little influence. For $k < j$ we specify the covariance[30] $E(\xi_j - \bar{\xi}_0)(\xi_k - \bar{\xi}_0)' = E(\xi_j - \bar{\xi}_0)(\xi_j - \bar{\xi}_0)'$. The small value for c_1 forces presample ξ_j to be close to ξ_k when j is close to k, again consistent with the observed month-to-month variation in $\hat{\xi}_{t|T}$.

Smoothed inferences about x_t using the full sample of available data, $\hat{x}_{t|T} = E(x_t|Y_T)$ and their variance matrix $P_{t|T} = E[(x_t - \hat{x}_{t|T})(x_t - \hat{x}_{t|T})']$ can be calculated by iterating backwards on the

---

[30] In other words,

$$P_{1|0} = \begin{bmatrix} c_0I_4 & c_0I_4 & c_0I_4 & \cdots & c_0I_4 \\ c_0I_4 & c_0I_4 + c_1I_4 & c_0I_4 + c_1I_4 & \cdots & c_0I_4 + c_1I_4 \\ c_0I_4 & c_0I_4 + c_1I_4 & c_0I_4 + 2c_1I_4 & \cdots & c_0I_4 + 2c_1I_4 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ c_0I_4 & c_0I_4 + c_1I_4 & c_0I_4 + 2c_1I_4 & \cdots & c_0I_4 + 47c_1I_4 \end{bmatrix}.$$

following equations for $t = T - 1, T - 2, ..., 1$:

$$J_t = P_{t|t} F' P_{t+1|t}^{-1}$$

$$\hat{x}_{t|T} = \hat{x}_{t|t} + J_t(\hat{x}_{t+1|T} - \hat{x}_{t+1|t})$$

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t'.$$

These smoothed inferences $\hat{x}_{t|T}$ and functions of them are plotted in Figures 5-7 and 10.

We calculated standard errors for the estimate $\hat{\theta}$ as in equation (3.13) in Hamilton (1994b):

$$E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' \simeq V = K_1^{-1} K_2 K_1^{-1}$$

$$K_1 = \left. \frac{\partial \ell(\theta)}{\partial \theta \partial \theta'} \right|_{\theta = \hat{\theta}}$$

$$K_2 = \sum_{t=1}^{T} \left\{ \left[\left. \frac{\partial \ln f(y_t|Y_{t-1}; \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} \right] \left[\left. \frac{\partial \ln f(y_t|Y_{t-1}; \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} \right]' \right\}.$$

To obtain standard errors for the variance decompositions in Figure 7 and Table 3, we generated $J = 1,000$ draws from the asymptotic distribution of $\hat{\theta}$, $\theta^{[j]} \sim N(\hat{\theta}, V)$, $j = 1, ..., J$ and calcuated $q_{s,k}(\theta^{[j]})$ as in equation (21) for each s and each $k = 1, ..., 4$. The standard deviation of $q_{s,k}(\theta^{[j]})/\sum_{k=1}^{4} q_{s,k}(\theta^{[j]})$ across draws j was used to get the error bands and standard errors in Figure 7 and Table 3.

The standard errors used for Figures 5 and 6 incorporate both filter and parameter uncertainty. The matrix $P_{t|T}$ summarizes uncertainty we would have about x_t even if we knew the true value of the parameters in θ. Given that we also have to estimate θ, the true uncertainty is greater than that represented by $P_{t|T}$. Following Ansley and Kohn (1986) we calculate the total variance as

$$P_{t|T}\big|_{\theta = \hat{\theta}} + Z_t V Z_t'$$

$$\underset{(4 \times 12)}{Z_t} = \left. \frac{\partial \hat{x}_{t|T}}{\partial \theta'} \right|_{\theta = \hat{\theta}}.$$

The values of $\{Z_t\}_{t=1}^{T}$ can be found by numerical differentiation, e.g., replace $\hat{\theta}$ with $\hat{\theta} + \delta e_i$ for $\delta = 10^{-8}$ and e_i the ith column of I_{12} and then redo the iteration to calculate $\hat{x}_{t|T}(\hat{\theta} + \delta e_i)$. The

ith column of Z_t is then $\delta^{-1}[\hat{x}_{t|T}(\hat{\theta} + \delta e_i) - \hat{x}_{t|T}|(\hat{\theta})]$.

C. Derivation of linearized variance and historical decompositions

The state equation $\xi_{t+1} = \xi_t + \varepsilon_{t+1}$ implies

$$
\begin{aligned}
\xi_{t+s} &= \xi_t + \varepsilon_{t+1} + \varepsilon_{t+2} + \varepsilon_{t+3} + \cdots + \varepsilon_{t+s} \\
&= \xi_t + u_{t+s}.
\end{aligned}
$$

Letting $y_t = (U_t^1, U_t^{2.3}, U_t^{4.6}, U_t^{7.12}, U_t^{13.+})'$ denote the (5×1) vector of observations for date t, our model implies that in the absence of measurement error y_t would equal $h(\xi_t, \xi_{t-1}, \xi_{t-2}, ..., \xi_{t-47})$ where $h(\cdot)$ is a known nonlinear function. Hence

$$
y_{t+s} = h(u_{t+s} + \xi_t, u_{t+s-1} + \xi_t, ..., u_{t+1} + \xi_t, \xi_t, \xi_{t-1}, ..., \xi_{t-47+s}).
$$

We can take a first-order Taylor expansion of this function around $u_{t+j} = 0$ for $j = 1, 2, ..., s$,

$$
y_{t+s} \simeq h(\xi_t, ..., \xi_t, \xi_t, \xi_{t-1}, ..., \xi_{t-47+s}) + \sum_{j=1}^{s} [H_j(\xi_t, \xi_t, ..., \xi_t, \xi_t, \xi_{t-1}, ..., \xi_{t-47+j})] u_{t+s+1-j}
$$

for $H_j(\cdot)$ the (5×4) matrix associated with the derivative of $h(\cdot)$ with respect to its jth argument. Using the definition of u_{t+j}, this can be rewritten as

$$
y_{t+s} \simeq c_s(\xi_t, \xi_{t-1}, ..., \xi_{t-47+s}) + \sum_{j=1}^{s} [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})] \varepsilon_{t+j}
$$

from which (18) follows immediately.

Similarly, for purposes of a historical decomposition note that the smoothed inferences satisfy

$$
\hat{\xi}_{t+s|T} = \hat{\xi}_{t|T} + \hat{\varepsilon}_{t+1|T} + \hat{\varepsilon}_{t+2|T} + \hat{\varepsilon}_{t+3|T} + \cdots + \hat{\varepsilon}_{t+s|T}
$$

where $\hat{\varepsilon}_{t+s|T} = \hat{\xi}_{t+s|T} - \hat{\xi}_{t+s-1|T}$. For any date $t+s$ we then have the following model-inferred value for the number of people unemployed:

$$\iota_5' h(\hat{\xi}_{t+s|T}, \hat{\xi}_{t+s-1|T}, \hat{\xi}_{t+s-2|T}, ..., \hat{\xi}_{t+s-47|T}).$$

For an episode starting at some date t, we can then calculate

$$\iota_5' h(\hat{\xi}_{t|T}, \hat{\xi}_{t|T}, \hat{\xi}_{t|T}, ..., \hat{\xi}_{t|T}, \hat{\xi}_{t-1|T}, ..., \hat{\xi}_{t+s-47|T}).$$

This represents the path that unemployment would have been expected to follow between t and $t+s$ as a result of initial conditions at time t if there were no new shocks between t and $t+s$. Given this path for unemployment that is implied by initial conditions, we can then isolate the contribution of each separate shock between t and $t+s$. Using the linearization in equation (18) allows us to represent the realized deviation from this path in terms of the contribution of individual historical shocks as in (22).

D. Alternative estimates of unemployment-continuation probabilities

There is an unresolved controversy in the literature about how to measure outflows from unemployment. Our measure described in footnote 1 follows van den Berg and van Ours (1996), van den Berg and van der Klaauw (2001), Elsby, Michaels and Solon (2009), Shimer (2012), and Elsby, Hobijn and Şahin (2013) in deriving flow estimates from the observed change in the number of unemployed by duration. An alternative approach, employed by Fujita and Ramey (2009) and Elsby, Hobijn and Şahin (2010), is to look at only those individuals for whom there is a matched observation of unemployment in month $t-1$ and a status of employment or out of the labor force in month t. In the absence of measurement error, the two estimates should be the same, but in practice they turn out to be quite different. One reason for the discrepancy is misclassification. For example, an individual who goes from long-term unemployed to out of the labor force to back to long-term unemployed in three successive months counts as a successful "graduate" from long-term unemployment using matched flows but is contributing to the stubborn persistence of long-term unemployment when using the stock data. A follow-up paper to Elsby, Hobijn and Şahin (2010) by Elsby et al. (2011) documented that of the individuals who were employed or out of the labor force

in month $t - 1$ and who were recorded as unemployed in month t, more than half reported their duration of unemployment to be 5 weeks or longer. Another important reason is that individuals for whom two consecutive observations are available differ in important ways from those for whom some observations are missing. Abowd and Zellner (1985) and Frazis et al. (2005) acknowledged that these measurement errors are more likely to bias the matched flow data than the stock data and suggested methods to correct the bias.

Since our goal is to understand how the reported stock of long-term unemployed came to be so high and why it falls so slowly, we feel that our approach, which is consistent with the observed stock data by construction, is preferable for our application.

E. Details of robustness tests

The standard errors in Table 3 were calculated as follows. For each model, we generated 500 draws for the k-dimensional parameter vector (where k is reported in the first row of the table) from a $N(\hat{\theta}, \hat{V})$ distribution where $\hat{\theta}$ is the MLE and $\hat{V}$ is the $(k \times k)$ variance matrix from inversed hessian of the likelihood function. For each draw of $\theta^{(\ell)}$ we calculated the values implied by that $\theta^{(\ell)}$ and then calculated the standard error of that inference across the draws $\theta^{(1)}, ..., \theta^{(500)}$.

Time-varying genuine duration dependence. Vishwanath (1989) and Blanchard and Diamond (1994) developed theoretical models in which genuine duration dependence could be linked to market tightness. See Whittaker and Isaacs (2014) for a detailed discussion of the conditions that can trigger extended unemployment benefits.

Shimer (2012) argued that this time-aggregation bias would result in underestimating the importance of outflows in accounting for cyclical variation in unemployment, and Fujita and Ramey (2009), Shimer (2012) and Hornstein (2012) all formulated their models in continuous time.

Allowing for structural shocks. For the factor model, the variance decomposition (19) becomes

$$
E(y_{t+s} - \hat{y}_{t+s|t})(y_{t+s} - \hat{y}_{t+s|t})'
$$

$$
= \sum_{j=1}^{s} [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})](\lambda\lambda' + Q)[\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]'
$$

$$
= \sum_{j=1}^{s} [\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]\lambda\lambda'[\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})]'
$$

$$
+ \sum_{j=1}^{s} \sum_{m=1}^{4} Q_m[\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})e_m][\Psi_{s,j}(\xi_t, \xi_{t-1}, ..., \xi_{t-47+j})e_m]'
$$

for Q_m the row m, column m element of Q.

Time aggregation. Elsby, Michaels and Solon (2009) questioned the theoretical suitability of a continuous-time conception of unemployment dynamics, asking if it makes any sense to count a worker who loses a job at 5:00 p.m. one day and starts a new job at 9:00 a.m. the next as if they had been unemployed at all. We agree, and think that defining the central object of interest to be the fraction of those newly unemployed in month t who are still unemployed in month $t + k$, as in our baseline model, is the most useful way to pose questions about unemployment dynamics. Nevertheless, and following Kaitz (1970), Perry (1972), Sider (1985), Baker (1992), and Elsby, Michaels and Solon (2009) we also estimated a version of our model formulated in terms of weekly frequencies as an additional check for robustness.

We can do so relatively easily if we make a few simplifying assumptions. We view each month t as consisting of 4 equally-spaced weeks and assume that in each of these weeks there is an inflow of w_{it} workers of type i, each of whom has a probability $p_{it}(0) = \exp[-\exp(x_{it})]$ of exiting unemployment the following week. This means that for those type i individuals who were newly unemployed during the first week of month t, $w_{it}[p_{it}(0)]^3$ are still unemployed as of the end of the month. Thus for the model interpreted in terms of weekly transitions, equations (10) and (11) would be replaced by

$$
U_t^1 = \sum_{i=H,L} \{w_{it} + w_{it}[p_{it}(0)] + w_{it}[p_{it}(0)]^2 + w_{it}[p_{it}(0)]^3\} + r_t^1
$$

$$
U_t^{2.3} = \sum_{i=H,L} \sum_{s=1}^{4} \{w_{i,t-1}[p_{i,t-1}(1)]^{8-s} + w_{i,t-2}[p_{i,t-2}(2)]^{12-s}\} + r_t^{2.3}
$$

for $p_{it}(\tau)$ given by (5) and (8). Note that although this formulation is conceptualized in terms of weekly inflow and outflows w_i and p_i, the observed data y_t are the same monthly series used in our other formulations, and the number of parameters is the same as for our baseline formulation.
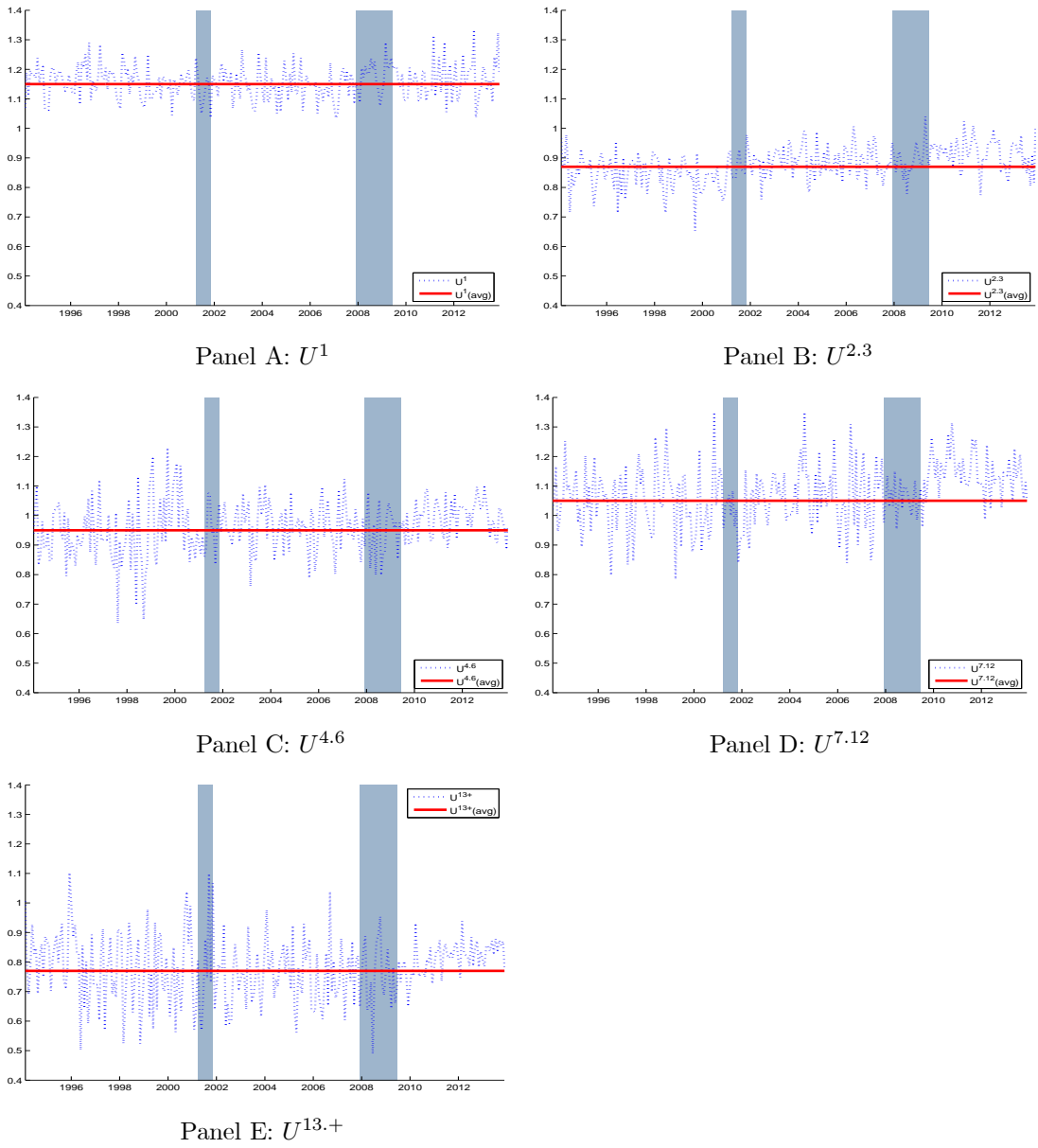
Panel A: U^1

Panel B: $U^{2.3}$

Panel C: $U^{4.6}$

Panel D: $U^{7.12}$

Panel E: $U^{13.+}$

Figure A1. Ratio of each duration group's share in the first and fifth rotation groups to that in all rotation groups

Appendix References

Abowd, John M., and Arnold Zellner (1985). "Estimating Gross Labor-Force Flows," *Journal of Business and Economic Statistics*, 3(3): 254-283.

Ansley, F. Craig, and Robert Kohn. (1986). "Prediction Mean Squared Error for State Space Models with Estimated Parameters," *Biometrika*, 73(2):467-473.

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

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

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin (2010). "The Labor Market in the Great Recession," *Brookings Papers on Economic Activity,* Spring 2010: 1-56.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin (2013). "Unemployment Dynamics in the OECD," *Review of Economics and Statistics*, 95(2): 530-548.

Elsby, Michael W. L., Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta (2011). "The Labor Market in the Great Recession: An Update to September 2011," *Brookings Papers on Economic Activity,* Fall 2011: 353-371.

Frazis, Harley J., Edwin L. Robison, Thomas D. Evans, and Martha A. Duff (2005). "Estimating Gross Flows Consistent with Stocks in the CPS," *Monthly Labor Review* September 2005: 3-9.

Hamilton, James D. (1994b). "State-Space Models," in Robert F. Engle and Daniel L. Mc-Fadden, eds., Handbook of Econometrics, Volume IV, Chapter 50, pp. 3039-3080. Amsterdam: Elsevier.

Kaitz, Hyman B. (1970). "Analyzing the Length of Spells of Unemployment," *Monthly Labor Review*, November 1970, pp.11-20, Bureau of Labor Statistics.

Perry, George L. (1972). "Unemployment Flows in the U.S. Labor Market," *Brookings Papers on Economic Activity,* 3(2):245–292.

Polivka, Anne E. and Stephen M. Miller (1998). "The CPS after the Redesign: Refocusing the Economic Lens," in John Haltiwanger, Marilyn E. Manser, and Robert Topel, eds., Labor Statistics Measurement Issues, pp.249-86.

Sider, Hal (1985). "Unemployment Duration and Incidence: 1968-1982," *American Economic Review*, 75(3):461-72.

Vishwanath, Tara. (1989). "Job Search, Stigma Effect, and Escape Rate from Unemployment," *Journal of Labor Economics*, 7(4):487-502.

Whittaker, Julie M. and Katelin P. Isaacs (2014). "Unemployment Insurance: Programs and Benefits," *Congressional Research Service Report for Congress*, 7-5700.