Measuring Labor-Force Participation and the Incidence and Duration of Unemployment

Hie Joo Ahn [*†] James D. Hamilton [‡]
Federal Reserve Board University of California, San Diego

March 16, 2019
Revised: March 10, 2021

Abstract

The underlying data from which the U.S. unemployment rate, labor-force participation rate, and duration of unemployment are calculated contain numerous internal contradictions. This paper catalogs these inconsistencies and proposes a unified reconciliation. We find that the usual statistics understate the unemployment rate and the labor-force participation rate by about two percentage points on average and that the bias in the latter has increased over time. The BLS estimate of the average duration of unemployment substantially overstates the true duration of uninterrupted spells of unemployment and misrepresents what happened to average durations during the Great Recession and its recovery.

Keywords: unemployment rate, labor-force participation rate, unemployment duration, measurement error

[*]The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. We thank Katharine Abraham, Alessandro Barbarino, Travis Berge, Michael Elsby, Andrew Figura, Glenn Follette, Ryan Michaels, Norm Morin, John Stevens, Robert Valletta, and anonymous referees for comments on earlier drafts of this paper and Jesse Wedewer and Michael Boerman for excellent research assistance. Data and software to reproduce results in this paper available at http://http://econweb.ucsd.edu/~jhamilton/AH2_code.zip.

[†]E-mail: hiejoo.ahn@frb.gov

[‡]E-mail: jhamilton@ucsd.edu

1 Introduction.

The Current Population Survey (CPS) is the primary source of information about the labor-force participation rate, unemployment rate, and duration of unemployment for the United States. There are multiple internal inconsistencies in the data from which the fundamental statistics are calculated– if one reported number is correct, another must be wrong. In this paper we catalog these inconsistencies and propose a unified reconciliation of all the problems.

Rotation bias. One source of inconsistency is rotation bias. In any given month, some households are being visited for the first time (rotation 1), others are being interviewed for the second time (rotation 2), with 8 different rotations contributing to the statistics reported for that month. One would think that in a random sample, the numbers calculated from different rotations for a given month should all be the same. But as documented by Hansen et al. (1955), Bailar (1975), Solon (1986), Halpern-Manners and Warren (2012) and Krueger, Mas, and Niu (2017), the reported unemployment rate differs significantly across rotations. In our sample (July 2001 to December 2020), the average unemployment rate among those being interviewed for the first time is 6.7%, whereas the average unemployment rate for the eighth rotation is 5.8%. Even more dramatic is the rotation bias in the labor-force participation rate. This averages 65.7% for rotation 1 and 64.0% for rotation 8 in our sample. Rotation bias affects any inference one draws from the CPS data. For example, it means that if one follows a fixed group of individuals over time, on average outflows from unemployment seem to exceed inflows.

Missing observations. A second source of inconsistency documented by Abowd and Zellner (1985) is that missing observations are not random. Meyer, Mok and Sullivan (2015) noted that households in the CPS have become increasingly less likely to answer surveys or to provide all answers. The standard approach is to calculate statistics for a given month based only on individuals for whom there is an observation that month. But if missing observations are not randomly drawn from the overall population, this may be an increasing source of bias in CPS estimates.

Reported job-search durations and observed continuation probabilities. A third problem in the CPS is inconsistency between the duration of job search reported by an individual in month t and the labor-force status recorded for that same individual in $t-1$. For example, consider those individuals who were counted as not in the labor force when in rotation 1 in month $t-1$ and unemployed when surveyed in rotation 2 in month t. In the second survey, the individual would be asked how long he or she has been looking for work. In an average month since 2001, two-thirds of these individuals' duration of unemployment is recorded as longer than 4 weeks and 17% of their durations are recorded as one or two years.

A related anomaly is the inconsistency between unemployment hazard rates and the reported duration of unemployment. For example, according to BLS adjusted numbers on labor-force flows, the average unemployed individual in 2011 had a 38% probability of exiting unemployment the following month. Among those already unemployed for more than 6 months, the probability was 31%.[1] From those probabilities we might expect an average duration of unemployment around (1/0.31) or 3 months. Yet according to the BLS, the average duration of unemployment among all those unemployed in 2011 was 40 weeks – three times the value that would be predicted on the basis of the reported hazards.

Number preference. A final source of inconsistency arises from people's preference for reporting certain numbers over others, as documented for example by Baker (1992), Torelli and Trivellato (1993), and Ryu and Slottje (2000). On average there are more people who say they have been looking for work for 6 months than say they have been looking for 23 weeks, though the fraction of those unemployed for 23 weeks should be greater than that of those unemployed for 6 months. In addition, people are more likely to report an even number of weeks than an odd number for shorter spells.

Our proposed resolution. Each of the problems above has been discussed in the literature. Previous papers addressed one problem in isolation. In this paper we show how these problems interact to influence the statistics that economists rely on and propose a unified resolution that addresses the issues simultaneously.

Our first step is to add a fourth category of labor-force status. We regard an individual in any month as either employed (E), unemployed (U), not in the labor force (N), or missing from the sample that month (M). On this basis we construct a data set in which all identities relating stocks and flows are respected; for example, the sum of EE, NE, ME, and UE transitions between $t-1$ and t exactly equals the total number of E at t. This has never been done before.

[1] The numbers cited here are described in Ilg (2011) which BLS provides on request. Our own direct estimates in Panel A of Figure 12 below suggest that hazards do not change much after durations beyond six months.

Our second step is to model statistically the way in which people's answers change the more times they have been interviewed. We interpret households in different rotations as being surveyed using a different interview technology and summarize how the differences in the average answers given by different rotations change gradually over time. We calculate the answer to the following counterfactual question: if a group of households in rotation j in month t were being interviewed for the first time instead of the jth time, how would their answers have been different?

Combining these two steps allows us to produce the first fully reconciled description of stocks and flows in the CPS data. By then looking at how reconciled ME, MN, and MU transitions differ from the rest of the population, we are able to adjust the treatment of missing observations based on what we know about those individuals when data are collected from them. We find that missing individuals are more likely than the general population to be unemployed. In addition, the biases introduced by missing observations have increased over time and are bigger when the labor market is slack. Our paper is the first to document the cyclical features in the bias coming from nonrandom missing observations.

With these tools we can provide the first-ever reconciled description of stocks and flows normalized on the basis of any of the eight interview technologies. In practice we need to choose a particular technology, which requires taking a stand on the source of the rotation bias. Much evidence, most of which is new to this paper, persuades us that the primary source of rotation bias is that individuals who are unemployed or only employed part-time perceive some stigma associated with those answers and become less engaged with the interview process the more times an interview is attempted. We conclude that the first-interview concept of unemployment is the best one to use, and show that the unemployment rate and labor-force participation rate are more seriously underestimated at points in time or for demographic groups for which the reported unemployment rate is higher.

Reconciling the inconsistencies in the CPS further requires confronting the gray area in the distinction between unemployed and not in the labor force. After examining the evidence we conclude that some of those currently counted as N should instead be regarded as actively seeking work. This turns out to be another reason why the unemployment rate and labor-force participation rate are more seriously underestimated at points in time or for demographic groups for which the unemployment rate is higher.

Finally, our resolution to the number-preference problem is to postulate a flexible latent distribution of perceived durations that is then reported by individuals with a certain structure of number-reporting preference; for related approaches see Baker (1992), Torelli and Trivellato (1993), and Ryu and Slottje (2000). Our approach is completely new compared to these studies in that our parameterization allows direct linkage of data on stocks, flows, and durations and in that both digit and interval preference are jointly considered. Our framework describes the reported values extremely accurately.

Why does it matter? The importance of these issues is illustrated in Panel A of Figure 1. This asks a very fundamental question: if someone is unemployed at $t-1$, what is the probability that person will still be unemployed at t? Researchers have used the CPS data to answer this question in two different ways. A measure based on reported unemployment durations calculates the ratio of individuals who are unemployed at t with a reported duration greater than 4 weeks to the total number of individuals unemployed at $t-1$. Variants of this calculation have been used by van den Berg and van der Klaauw (2001), Elsby, Michaels and Solon (2009) and Shimer (2012). This measure is plotted as the black line in Panel A. An alternative measure based on labor-force flows looks at the subset of individuals who are U at $t-1$ and either E, N, or U at t and calculates the number of UU continuations as a fraction of the sum. Variants of this approach were used by Fujita and Ramey (2009) and Elsby, Hobijn and Şahin (2010). The flow-based measure is plotted as the green line. If all magnitudes were measured accurately the two estimates should give a similar answer. But in practice they are wildly different. The duration-based measure averages 70.3% over our sample, while the flow-based measure averages 53.0%.

These differences are caused by the multiple inconsistencies mentioned above. The flow-based measure underestimates the true continuation probability because (1) some UN transitions are a result of rotation bias and (2) some UN transitions should be interpreted as UU continuations. The duration-based measure overestimates the probability, because a substantial number of people interpret the duration of job search as including on-the-job search or the time since the last salient job; see Elsby et al. (2011), Farber and Valletta (2015), and Kudlyak and Lange (2018). Indeed, in April 2020 the duration-based approach would calculate an unemployment-continuation probability of 122%! This is because the number of people unemployed reporting 5-week or longer durations in April was 22% higher than the total number of people counted as unemployed in March. Our reconciled estimate is shown in the blue line in Panel A. It usually falls in between the other two estimates, but agrees with the flows-based conclusion that in the spring of 2020 many people left the labor force rather than continuing to look for work.

Another fundamental question is, how many people become unemployed each month? One estimate (e.g., Shimer, 2012) is simply the number of unemployed individuals reporting durations of less than 5 weeks. The black line in Panel B of Figure 1 shows this value as a percent of the civilian noninstitutional population. As noted by Elsby et al. (2011), it underestimates new inflows into unemployment since half of *EU* and *NU* transitions report unemployment durations of 5 weeks or longer. Alternatively, the BLS publishes separate estimates of *EU* and *NU* flows that they adjust to address some of the problems that we document in this paper. However, our analysis suggests that their adjustments are incomplete. Our reconciled series (in blue) is often significantly higher than the BLS adjusted estimate (in turquoise).

Panel C of Figure 1 compares our adjusted estimate of the unemployment rate with the BLS estimate. Our measure is 2.0% higher on average, and the gap increased during both the Great Recession and the pandemic recession. The gap between our measure and the BLS measure of the labor-force participation rate (Panel D) is 2.1% on average, and this gap also increased in both recessions. Our estimates imply that the labor-force participation rate declined by 1.1% less between 2001 and 2020 than is implied by the BLS series.

Whereas BLS estimates of unemployment duration are based on individuals' reported durations of job search, our estimates are based on uninterrupted spells of unemployment. In going from the green to blue lines in Panel A, we adjusted unemployment continuations up considerably from the standard estimates, but we did not adjust these all the way up to those implied by reported durations in black. As a result, our reconciled estimates of average unemployment durations (shown as blue in Panel E) are considerably below those from BLS (black), similar to the conclusion by Kudlyak and Lange (2018). Our estimates of average duration did not rise as much during the Great Recession as suggested by the BLS series based on reported durations. Also, our reconciled estimates subsequently recovered to pre-recession levels, whereas the BLS reported durations do not.

A significant part of the measurement errors we discuss arises from ambiguities in classifying individuals as "unemployed" versus "not in the labor force." The employment-to-population ratio

(Panel F) avoids these issues and thus might be a better indicator of labor market slack. However, the employment-to-population ratio is still influenced by rotation bias, which we attribute to stigma and disengagement with repeated interviews of some part-time workers.

The plan of the paper is as follows. Section 2 describes the structure of the CPS survey and how we construct a unique data set in which the accounting identities that should relate flows to changes in stocks all hold. Section 3 characterizes rotation bias and describes our solution, while Section 4 does the same for nonrandomly missing observations. Section 5 investigates individuals who report having been searching for a job longer than is consistent with their recorded labor-force-status history. Section 6 brings all the elements together to develop our final reconciliation of stocks, flows, and unemployment durations. Section 7 compares our adjustments with others that have been proposed, while Section 8 concludes.

2 Constructing a data set in which accounting identities all hold.

The first step in producing reconciled estimates is to construct a data set in which the accounting identities that should relate flows to changes in stocks all hold. Since July 2001, each month around 60,000 housing units are included in the Current Population Survey. An effort is made to contact each address and determine the number of individuals aged 16 or over who are not in the armed forces or in an institution such as prison or a nursing home. An individual is counted as employed (E) if during the reference week of the survey month the individual did any work at all for pay, for their own business, or were temporarily absent from work due to factors like vacations, illness, or weather. People are counted as unemployed (U) if they were not E but were available for work and made specific efforts to find employment some time during the previous 4 weeks or on temporary layoff from their work. Individuals who are neither E nor U are counted as not in the labor force (N). One person in the household can provide separate answers for each of the individuals living at that address.

The next month and each of the following two months, an interviewer attempts to contact the same address to ask the same questions. In any given month, around 1/8 of the 60,000 qualifying households are being interviewed for the first time (denoted rotation 1), and another 1/8 each are being interviewed for the second, third or fourth time (rotations 2, 3, or 4). After the fourth month the household is not interviewed for the next 8 months, but is reinterviewed again 1 year after the first interview (rotation 5) and again for each of the following 3 months (rotations 6, 7, and 8). For data since 1994, if an individual was unemployed in two consecutive months, the interviewer does not ask again the duration of unemployment the second month, but simply adds time elapsed since the previous interview to the previous answer. Thus new unemployment duration data are only collected in rotations 1 and 5, or in the other rotations for someone who was E, N or missing from the sample the month before.

The survey is imperfect for purposes of tracking the experience of an individual across months. The CPS micro database contains identifiers that in principle can uniquely identify each individual.[2] However, as noted by Feng (2001), Madrian and Lefgren (2000), and Nekarda (2009), significant clerical errors can arise for example from switching the order in which individual household members are listed. We follow researchers like Madrian and Lefgren (2000), Nekarda (2009), and Shimer (2012) in regarding the status of an individual from the previous rotation as being missed in the current rotation if fundamental characteristics of the individual don't match. Specifically, we treat data in two different rotations as representing different individuals if the gender is not the same or if the age (after top-coding at age 65) differs by more than two years. This is a more conservative criteria than used by other researchers, who often also insist on consistency in reported race or education, attributes which themselves may be prone to reporting error. Unlike other researchers, we keep track of missing individuals and note if they re-enter the sample in a later rotation.

The raw data for our study thus consist of $y_{X,t}^{[j]}$, the sum of the number of individuals (multiplied by a weight associated with that individual) who are in rotation $j \in \{1, ..., 8\}$ in month t with reported status $X \in \{E, N, M, U\}$, and $y_{X_1,X_2,t}^{[j]}$, the weighted sum of individuals reporting X_1 in rotation $j-1$ in month $t-1$ and X_2 in rotation j in month t for $j \in J = \{2,3,4\} \cup \{6,7,8\}$. See Table A-1 in the online appendix for a summary of notation used in this study. A key advantage of our approach is that, unlike the values used by any other researchers, our data on stocks and flows are internally consistent by construction, always satisfying the accounting identities

$$y_{X_2,t}^{[j]} = y_{E,X_2,t}^{[j]} + y_{N,X_2,t}^{[j]} + y_{M,X_2,t}^{[j]} + y_{U,X_2,t}^{[j]} \quad (1)$$

[2] The values of HRHHID and HRHHID2 together uniquely identify the household and then PULINENO should identify the individual within the household.

$$y^{[j-1]}_{X_1,t-1} = y^{[j]}_{X_1,E,t} + y^{[j]}_{X_1,N,t} + y^{[j]}_{X_1,M,t} + y^{[j]}_{X_1,U,t} \tag{2}$$

for all t, X_1, X_2 and $j \in J$.

Note that our approach generates a complete history for everyone in the middle of the sample, and includes for example paths like $EEME{-}EMEE$. However, we need 16 months of observations to construct such records, and this creates problems at the start and end points of the sample. For example, our sample starts in 2001:7. Consider someone whose history if followed for 16 months would have been $EEMM - MMMM$. If that history began in 2001:7, the individual would be included in our database. But if the same history had begun in 2001:5, the individual would never appear in our sample.[3] This causes the number of individuals who are classified as M to be artificially depressed in the first year of the sample. A similar effect arises at the end of the sample, with individuals whose record would have been $MMEE - EEEE$ not being apparent if their rotations 1 or 2 would have come at the end of the sample. We therefore adjusted the counts of M and MM at the beginning and end of the sample upward based on the average counts of M for each rotation over the nearest year of complete observations; for details see Appendix A. Since changes in M occur relatively slowly in our sample, this adjustment has little effect on any of the key measures we develop. We made additional adjustments when new households were added and other households dropped in the 2004 and 2014 sample redesigns.[4]

BLS also assigns a weight to each individual. People with characteristics that are underrepresented in a particular month are given a larger weight. These weights are a partial response of BLS to the issue that missing individuals are not a random sample of the population. We want to include this correction to demonstrate the need for additional corrections for missing individuals. We can not use the exact BLS weights to do this because the BLS may assign a given individual different weights in two different months, which is another reason in addition to missing observations why (1) and (2) do not hold in the BLS data. Our approach was to assign a fixed weight for an individual across all 8 possible observations based on the BLS weight for that individual in the first month for which data are recorded for that person, as described in Appendix A.

[3] See Appendix A for detailed examples.

[4] With the expansion of the survey from 50,000 to 60,000 households, beginning in July 2001, some individuals were added and others dropped across a number of rotations, with waves of new individuals added to subsequent rotations 5. Tracking individuals before and after this break is considerably harder than handling the sample redesign in 2004 and 2014. For this reason we simply begin our analysis with the modern design adopted in July 2001.

3 Rotation bias.

In this section we summarize how people's answers to questions change the more times they are asked and then develop our solution.

3.1 Key facts about rotation bias.

Table 1 reports the monthly average number of sampled individuals with measured labor force status E, N, M, or U for each of the 8 rotation groups.[5] Column 6 shows that the average unemployment rate declines sharply as a function of rotation group, starting out at 6.7% for rotation 1 but falling all the way to 5.8% for rotation 8. Column 7 reveals another interesting fact that has not been much commented on in the earlier literature: the measured labor-force participation rate falls even more sharply. Column 3 documents a third tendency– individuals are much more likely to be missed in rotation 1 and 5 compared to other groups.

We summarize these tendencies with some simple regressions. Let $x_t^{[j]} = 100 y_{X,t}^{[j]} / \left(y_{E,t}^{[j]} + y_{N,t}^{[j]} + y_{M,t}^{[j]} + y_{U,t}^{[j]} \right)$ denote the percentage of individuals in rotation group j sampled in month t with measured status $X = E, N, M$, or U; thus $e_t^{[j]} + n_t^{[j]} + m_t^{[j]} + u_t^{[j]}$ exactly equals 100 for every j and every t. Consider an 8-variable panel regression with time fixed effects where the dependent variable is $n_t^{[j]}$, $j = 1, ..., 8, t = 1, ..., T$:

$$n_t^{[j]} = \alpha_{nt} + \delta_n j + \alpha_{n1} d_{1t} + \alpha_{n5} d_{5t} + \varepsilon_{nt}^{[j]}. \tag{3}$$

Here α_{nt} is the time fixed effect for month t, δ_n captures a linear trend across rotations (with increased fraction of N in later rotations captured by $\delta_n > 0$), $d_{1t} = 1$ if $j = 1$ and 0 otherwise allows for something special about the first rotation group, while $d_{5t} = 1$ if $j = 5$ serves a similar function for rotation 5. The fitted value of this regression (with fixed effect $a_{nt} = 0$) is plotted as the red curve in Figure 2. These coefficients capture the tendency for the percentage of individuals classified as N to increase sharply across rotation groups.

Coefficients for panel regressions in which $e_t^{[1]}$,...,$e_t^{[8]}$ are the 8 dependent variables are plotted as the thick black curve in Figure 2. Coefficients when unemployment is the dependent variable

[5] For example, the entry in the first row and column is $T^{-1} \sum_{t=1}^{T} y_{E,t}^{[1]}$.

are plotted as the dashed green line. The rising trend across rotations in N ($\delta_N = 0.0010$) is accounted for by falling trends in E and U ($\delta_E + \delta_U = -0.0012$). The bulges in M in rotation 1 ($\alpha_{M1} = 0.0168$) and rotation 5 ($\alpha_{M5} = 0.0153$) are accounted for by drops in E and N in those rotations.[6]

A given household thus often gives different answers depending on the number of times the household has previously been interviewed. We interpret this as differences in interview technology: the process by which data are obtained differs across rotations, and the numbers from different rotations mean different things. As a first step we summarize these differences in the form of a counterfactual question: if an individual in rotation j had instead been interviewed using the technology i, how would their answers have differed? In Section 3.2 we show how to answer this question for $i = 1$ and then find the answer for any i. Section 3.3 discusses which interview technology i should be used as a baseline summary of the data. We identify several reasons why we prefer to use the answers that people give the first time they are interviewed ($i = 1$).

3.2 Adjusting stocks and flows for rotation bias.

Let $\pi_t^{[j]} = (\pi_{E,t}^{[j]}, \pi_{N,t}^{[j]}, \pi_{M,t}^{[j]}, \pi_{U,t}^{[j]})'$ denote the observed fraction of individuals who reported status X when interviewed in rotation j in month t. For each $j \in J = \{2,3,4\} \cup \{6,7,8\}$, of the individuals who reported status X_1 in rotation $j-1$ in month $t-1$, some fraction $\pi^j_{X_1,X_2,t}$ are observed to report status X_2 in rotation j for $X_i \in \{E, N, U, M\}$; thus $\pi_{XE,t}^{[j]} + \pi_{XN,t}^{[j]} + \pi_{XU,t}^{[j]} + \pi_{XM,t}^{[j]} = 1$ for all X, t and $j \in J$. Collect these observed probabilities in a matrix

$$\Pi_t^{[j]} = \begin{bmatrix} \pi_{EE,t}^{[j]} & \pi_{NE,t}^{[j]} & \pi_{ME,}^{[j]} & \pi_{UE,t}^{[j]} \\ \pi_{EN,t}^{[j]} & \pi_{NN,t}^{[j]} & \pi_{MN,t}^{[j]} & \pi_{UN,t}^{[j]} \\ \pi_{EM,t}^{[j]} & \pi_{NM,t}^{[j]} & \pi_{MM,t}^{[j]} & \pi_{UM,t}^{[j]} \\ \pi_{EU,t}^{[j]} & \pi_{NU,t}^{[j]} & \pi_{MU,t}^{[j]} & \pi_{UU,t}^{[j]} \end{bmatrix} \qquad j \in J.$$

[6] These findings are consistent with Krueger, Mas, and Niu's (2017) finding that rotation-group bias is associated with nonresponses and with Bailar's (1975) conclusion that the rotation-group bias of the unemployment rate can be explained by the participation margin.

Notice that each column of $\Pi_t^{[j]}$ sums to unity. Our constructed data set exactly satisfies the accounting identity

$$\Pi_t^{[j]}\pi_{t-1}^{[j-1]} = \pi_t^{[j]} \quad \text{for all } t \text{ and } j \in J. \tag{4}$$

The rotation-bias problem is that the transition matrix $\Pi_t^{[j]}$ is measured in units of going from interview $j-1$ to interview j. We would like to describe transition probabilities using a fixed-interview technology, which we denote π_t^*:

$$\Pi_t^*\pi_{t-1}^* = \pi_t^*. \tag{5}$$

In this section we describe how to use observations on $\pi_t^{[j]}$ and $\Pi_t^{[j]}$ to construct values of π_t^* and Π_t^* satisfying (5). To do this, we need to choose units in which π_t^* is to be measured. We first illustrate the solution when the first-interview technology is used and then show how the same ideas can be adapted to normalize in terms of any interview technology.

We can summarize how interview j answers in month t are observed to differ from interview 1 answers in terms of a (4×4) matrix $R_t^{[j]}$:

$$R_t^{[j]}\pi_t^{[j]} = \pi_t^{[1]} \qquad \text{for } t = 1, ..., T \text{ and } j = 2, ..., 8. \tag{6}$$

This can be done in many different ways. Since elements of both $\pi_t^{[j]}$ and $\pi_t^{[1]}$ sum to unity, choosing 3 parameters in $R_t^{[j]}$ is all that is necessary to ensure that (6) holds as an exact accounting identify for every j and every t. We choose 3 magnitudes that are based on the differences between rotation j and rotation 1 in the fractions reported to be employed, unemployed, and not in the labor force.[7] Since elements of $\pi_t^{[j]}$ sum to unity, these three numbers also completely describe differences between rotations in the fractions that are M.

Consider taking equation (4) for $j = 2$ and premultiplying both sides by $R_t^{[2]}$:

$$R_t^{[2]}\Pi_t^{[2]}\pi_{t-1}^{[1]} = R_t^{[2]}\pi_t^{[2]}. \tag{7}$$

From (6), the right side of (7) equals $\pi_t^{[1]}$, which could be regarded as an estimate of π_t^* if we were

[7] Specifically, $\theta_{EM,t}^{[j]} = 1 - \pi_{E,t}^{[1]}/\pi_{E,t}^{[j]}$, $\theta_{NU,t}^{[j]} = (\pi_{U,t}^{[1]} - \pi_{U,t}^{[j]})/\pi_{N,t}^{[j]}$, and $\theta_{NM,t}^{[j]} = 1 - \theta_{NU,t}^{[j]} - \pi_{N,t}^{[1]}/\pi_{N,t}^{[j]}$. For details see Appendix E.

normalizing on the first-interview technology. The left side is of the form $\Pi_t^* \pi_{t-1}^*$ for $\Pi_t^* = R_t^{[2]} \Pi_t^{[2]}$. Thus (7) is of the desired form (5): premultiplying $\Pi_t^{[2]}$ by $R_t^{[2]}$ achieves the goal of expressing the values for rotation 2 in month t that are implied by $\Pi_t^{[2]}$ into units of rotation 1 answers. More generally, if we premultiply (4) by $R_t^{[j]}$,

$$R_t^{[j]} \Pi_t^{[j]} (R_{t-1}^{[j-1]})^{-1} R_{t-1}^{[j-1]} \pi_{t-1}^{[j-1]} = R_t^{[j]} \pi_t^{[j]}$$

$$R_t^{[j]} \Pi_t^{[j]} (R_{t-1}^{[j-1]})^{-1} \pi_{t-1}^{[1]} = \pi_t^{[1]} \qquad \text{for } j \in J. \tag{8}$$

This gives another expression of the form of (5) with now $\Pi_t^* = R_t^{[j]} \Pi_t^{[j]} (R_t^{[j-1]})^{-1}$. In this case, we translate both the interview j technology to which $\Pi_t^{[j]}$ goes and the interview $j-1$ technology from which it comes into interview 1 answers.

Thus a single proposed value for Π_t^* implies predicted values for the 6 observed transition matrices $\Pi_t^{[j]}$ for $j \in \{2, 3, 4, 6, 7, 8\}$. If we have an estimate of π_{t-1}^*, the value of Π_t^* also predicts that we should observe $\pi_t^{[1]} = \Pi_t^* \pi_{t-1}^*$ and further predicts $\pi_t^{[5]} = (R_t^{[5]})^{-1} \Pi_t^* \pi_{t-1}^*$. Our approach is to choose as the estimate of Π_t^* the value that minimizes the sum of squared errors of all these predictions. For details see Appendix E.

Estimates based on full-sample averages. We first apply this approach to the full-sample average values of $\pi^{[j]}$ and $\Pi^{[j]}$. The full-sample estimates of the three parameters in $R^{[j]}$ are reported in Table 2. The first row reports that on average there are 1-2% more individuals employed in rotations 2-4 or 6-8 compared to rotation 1. This primarily reflects working people who are missed in the first interview. On the other hand, rotation 5 (which follows an 8-month break) reports similar numbers of E as rotation 1 ($\theta_{EM}^{[5]}$ near 0).[8] The second row captures a rising tendency for those who would have been counted as N in later rotations to have been counted as U in the first interview. The third row indicates that a large and rising fraction of those counted N in later rotations would have been M in rotation 1.

The full-sample estimates of π^* and Π^* are reported in Table 3. The value of $\pi^{[j]}$ that would be predicted by our approach is compared with the observed values for each rotation in Figure

[8] The estimate of $\theta_{EM}^{[5]}$ is actually very slightly negative (-0.0049). The value reported in Table 2 and used in the calculations below sets $\theta_{EM}^{[5]} = 0$ if the value implied by (A9) is negative. This makes essentially no difference for any results.

3. Our representation fits the values of each $\pi^{[j]}$ essentially perfectly. The model's prediction for the full-sample average value of $\Pi^{[j]}$ is compared with the observed values in Figure 4 for $j \in J$.[9] These show a reasonable fit, though not perfect. One could try to model in more detail features such as the tendency for those missing in rotation 1 to be reported as employed in rotation 2 and for those not in the labor force in rotation 1 to be missing in rotation 2. Notwithstanding, our simple framework does a reasonable job of capturing transitions.

Estimates using alternative interview technologies. We illustrated how to calculate the value of π^* in terms of the rotation 1 technology. But now that we've found π^*, we can easily calculate the answer using any other technology. For example, $(R^{[5]})^{-1}\pi^*$ gives the answer in terms of the rotation 5 technology. The BLS approach, which simply averages the rotations together, is implicitly reporting results in terms of an average technology, which in our formulation would be described as $\pi^{**} = \tilde{R}^{-1}\pi^*$ for $\tilde{R}^{-1} = (1/8)\sum_{j=1}^{8}(R^{[j]})^{-1}$. Appendix Table A-5 reports π^{**} and $\Pi^{**} = \tilde{R}^{-1}\Pi^*\tilde{R}$, our estimates of the full-sample averages and transition probabilities if all individuals had been surveyed using the average interview technology.

Month-by-month estimates. To construct month-by-month estimates of π_t^* and Π_t^*, we take the view that the nature of rotation bias changes only gradually over time. We implemented this using the idea of exponential smoothing, replacing counts for each rotation $y_{Xt}^{[j]}$ with a weighted average of recent counts for that rotation:

$$\overline{y}_{X,t}^{[j]} = (1-\lambda)y_{X,t}^{[j]} + \lambda\overline{y}_{X,t-1}^{[j]}. \tag{9}$$

For $\lambda = 1$, this method would reproduce the full-sample averages just reported. For $\lambda = 0$, it would amount to estimating values for each month in isolation of all the others. We set $\lambda = 0.98$, which means that observations 3 years prior to t receive about one-half the weight of observation t in contributing to the smoothed count $\overline{y}_{X,t}^{[j]}$.[10] We then calculated fractions $\overline{\pi}_t^{[j]}$ from these smoothed counts, and smoothed the resulting $\theta_t^{[j]}$ as well.

The resulting series for $\overline{\theta}_{EM,t}^{[j]}$, $\overline{\theta}_{NU,t}^{[j]}$, and $\overline{\theta}_{NM,t}^{[j]}$ are plotted in Figure 5. The value of $\overline{\theta}_{EM,t}^{[j]}$,

---

[9] Note we do not offer a predicted value for transitions from $X^{[4]}$ to $X^{[5]}$ since there are 8 intervening months between rotations 4 and 5.

[10] That is, $0.98^{36} = 0.48$. We started the recursion by setting $\bar{y}_{X,1}^{[j]} = (1/12)\sum_{t=1}^{12} y_{X,t}^{[j]}$ the average of the first year of observations.

which characterizes the tendency to record people as E in rotation j who would have been M in rotation 1, has fallen somewhat over time. By contrast, $\overline{\theta}_{NU,t}^{[j]}$, which governs the tendency of people who would have been counted as U in earlier rotations to be designated as N in later rotations, has increased over time. The third parameter, $\overline{\theta}_{NM,t}^{[j]}$, which characterizes the tendency of someone who would have been counted as M in rotation 1 to be counted as N in later rotations, has not changed much over time.

We used these smoothed estimates to construct $\bar{R}_t^{[j]}$ and found π_t^* as described in Appendix E for each t. Note that the estimate π_t^* is a function primarily of the raw data $\pi_t^{[j]}$ for month t alone. Smoothing is used solely for the purpose of converting $\pi_t^{[j]}$ counts into $\pi_t^{[1]}$ counts based on the recent historical average relation between $\pi_t^{[j]}$ and $\pi_t^{[1]}$. Note also that the estimate π_t^* is based on the combined values of $\pi_t^{[j]}$ across all rotations j, just as the BLS estimates for month t are. The difference is that whereas BLS estimates are based on averaging of $\pi_t^{[j]}$ across rotations j, our estimate combines the different $\pi_t^{[j]}$ in a way that takes into account the systematic differences between answers given in different rotations. This estimate π_t^* is the starting point for constructing fully adjusted estimates described below.

3.3 Choosing a baseline interview technology.

The framework in the previous subsection allows us to reconcile stocks and flows in the CPS data and summarize that reconciliation using any interview technology. In practice we need to choose a particular technology as a baseline. In this section we review the reasons why we recommend using the first-interview definition of labor-force status.

Disengagement. The tendency to report a higher incidence of unemployment the first time people are asked has also been observed in the Netherlands (van den Brakel and Krieg, 2015) and New Zealand (Silverstone and Bell, 2010). One possible explanation is that people become less engaged the more times they are interviewed and tend toward answers that they think will end the interview more quickly. For example, the CPS interview is more onerous if the respondent says that they have worked at more than one job. The number of people reporting more than one job drops sharply across rotations (Halpern-Manners and Warren, 2012; Hirsch and Winters, 2016). The CPS questionnaire also routes people over age 50 who say they are retired through an abbreviated set of labor-force questions.[11] It is interesting to note that more than all of the increasing incidence of N in later rotations can be explained by larger numbers of people who say they are retired or disabled. Specifically, note from row 1 in Table 4 that the average fraction of the population categorized as N is 1.3% higher in rotations 2-8 than in rotation 1. Row 2 shows that the fraction of the population categorized as N and retired is 1.0% higher in 2-8, and row 3 shows that the fraction categorized as N and disabled is 0.6% higher.

This raises the possibility that some of the people who had reported U in rotation 1 hoped to end the interview more quickly in later interviews if they claimed to be retired or disabled. We can observe in the data that those who are allegedly retired or disabled in rotations 2-7 are more likely to return to the labor force (that is, to report E or U) the following month than are the retired or disabled in rotation 1 (see row 4 of Table 4). This observation is consistent with the inference that some of the additional individuals in later rotations who are designated as N are in an objective sense still in the labor force.

Stigma. Another possible explanation suggested by Halpern-Manners and Warren (2012) is that some people may perceive a stigma in reporting to an official government agency that they are continually searching for a job without success. This of course could interact with the disengagement effect – someone who feels some stigma associated with their status may become less engaged with the interview process than the general population. This could lead some respondents to report in subsequent interviews that they did not actively search for work even though they did, which would show up as an increase in N and decrease in U in later rotations. The CPS allows one member of the household to report the labor-force status for all the adults living there. It is noteworthy that unemployment falls much more quickly across rotations among individuals who are reporting their own status compared to individuals whose status is reported by a proxy, consistent with Halpern-Manners and Warren's hypothesis. Self-responders account for half of the total observations but two-thirds of rotation bias (see rows 6 and 7 of Table 4).

Confusion or learning. An alternative hypothesis is that some individuals are confused by the questions or learn the meaning better as interviews are repeated. Some interesting evidence on this comes from differences in rotation bias across demographic groups and over time. Following Krueger, Mas and Niu (2017), we summarize the magnitude of rotation bias by the slope of a

[11] Current Population Survey Interviewing Manual, April 2015, p. B3-3.

regression of the unemployment rate for rotation j on a constant and the month-in-sample j. For the subsample 2001:7 to 2008:6, the slope of this regression is -0.11, meaning that on average, rotation j reported an unemployment rate that is 0.11 percentage points lower than that reported by rotation $j-1$. For those without a high school degree the slope was -0.21 while for those with a college degree the slope was only -0.05. These differences might seem to lend support to the confusion/learning hypothesis. Nonetheless, we note that even for those with college degrees, the slope is still highly statistically significant, with a standard error below 0.01.

Another feature that distinguishes less educated workers is that they have higher unemployment rates than the general population. The top panel of Figure 6 illustrates the importance of this graphically. The horizontal axis plots the average unemployment rate as measured by BLS for a particular demographic group and sample period. The vertical axis plots the absolute value of the slope from the Krueger, Mas and Niu regression. For the overall population, the average unemployment rate over 2001-2008 was 5.3%, represented by the point A1 $= (5.3, 0.11)$ in the figure. College graduates (represented by the point C1) had an average unemployment rate over this period of 2.6%, about half that of the overall population, just as the slope for this group is about half that for the overall population for this period. The average unemployment rate for those without a high school diploma, 11.2%, is about twice that for the overall population, as is the slope for this group, as represented by point L1 in the figure.

The variable whose value is plotted on the vertical axis in the top panel of Figure 6 is different for different demographic groups and different sample periods. Consider a second sample, 2008:7-2014:6, that includes the high unemployment rates in the aftermath of the Great Recession. For almost every demographic group, the unemployment rate is higher in the second subsample than in the first, and the magnitude of rotation bias goes up by a roughly proportionate amount. It seems implausible that the college-educated (S2 and C2) were more confused by the questions during the Great Recession than they had been in the earlier decade. Nor is the increase in rotation bias part of a long-term trend. Over the 2014:7-2020:1 subsample, we see slopes and unemployment rates come back down together for every group.

We can summarize this regularity with an OLS regression fit to the 27 data points in the top panel of Figure 6 (standard errors in parentheses):

$$s_{iT} = \underset{(0.010)}{0.029} + \underset{(0.0013)}{0.0126} u_{iT} + \hat{\varepsilon}_{iT} \qquad R^2 = 0.80. \tag{10}$$

Here s_{iT} is the negative of the Krueger-Mas-Niu slope coefficient[12] for demographic group i and subsample T and u_{iT} is the average unemployment rate for that group. Most of the variation in the slope across demographic groups and across time can be explained by differences in the unemployment rate across groups and across time. The data are suggestive of a universal law: if a percentage u_{it} of a group i at date t are truly unemployed, 1.26% of those unemployed individuals will no longer be classified as unemployed in each subsequent interview. This law seems to hold for every education group, demographic group, and point in time. The regression means that rotation bias can be explained statistically solely on the basis of the underlying true unemployment rate and has nothing to do with education or age. This observation casts doubt on the hypothesis that rotation bias arises from confusion or learning about the questions and is exactly what we would expect to see if stigma and disengagement are the main explanation for rotation bias.

Rotation bias in the employment to population ratio. We observe differences across rotations not just in the unemployment rate but also in the employment rate. The employment-to-population ratio is 0.9 percentage points lower on average for rotations 2-8 compared to rotation 1 (see row 8 of Table 4). Denote by $\dot{E}_t$ the subset of individuals who worked only part-time in month t or who are usually employed but didn't work in t. Row 9 shows that rotation bias in E can be entirely explained by the rotation bias in $\dot{E}$. It is possible that some people are ashamed of only being able to find part-time work and for this reason also become less engaged in subsequent interviews.

The bottom panel of Figure 6 reproduces the analysis in the top panel with U now replaced by $\dot{E}$. Here $\dot{e}_{it}^{[j]}$ is the fraction of people in rotation j, demographic group i and month t who only work part time in t or report that they are usually employed but not in month t. Let $\dot{s}_{iT}$ be the negative of the slope coefficient from a regression of $\dot{e}_{it}^{[j]}$ on a constant and j over subsample T. The bottom panel of Figure 6 shows the scatterplot relating $\dot{s}_{iT}$ to the average value of $\dot{e}_{it}$ for that

[12] That is, $s_{iT} = -\hat{\beta}_{iT}$ in the regression $u_{it}^{[j]} = \alpha_{iT} + \beta_{iT} \cdot j + \varepsilon_{it}$ for $t \in T$.

group and subsample. A regression line fit to these 27 data points is

$$\dot{s}_{iT} = \underset{(0.035)}{0.014} + \underset{(0.0021)}{0.0077}\dot{e}_{iT} + \hat{\varepsilon}_{iT} \qquad R^2 = 0.34. \tag{11}$$

This does not have as good a fit as (10) – demographics and sample period play a more important role in rotation bias in the unemployment rate than in the employment-to-population ratio. Notwithstanding, the slope coefficient in (11) is highly statistically significant.

College-educated individuals are significant outliers in the second panel of Figure 6. Part-time work for these individuals may involve consulting or programming for which they may feel little or no stigma compared to less-educated individuals whose part-time work may be cleaning or serving. In fact, the slope coefficients corresponding to the height of C1, C2 and C3 are each far from statistically significant. When these three observations are dropped, the R^2 rises to 0.49.

Evidence from reported durations of job search. One of the main objectives of our study is to reconcile the discrepancies between different CPS statistics. One important inconsistency is that reported durations of unemployment are much longer than could be consistent with observed probabilities of UU continuations. Because more U get counted as N as we increase the number of interviews j, if we were to reconcile stocks and flows on the basis of the interview $j-1$ technology, some of the observed UN transitions between rotation $j-1$ and j would be interpreted as UU continuations, decreasing the inconsistency between UU continuations and reported durations. By contrast, if we were to standardize on the basis of interview j technology, some of the reported UN transitions would be interpreted as NN continuations, increasing the inconsistency between reported durations and the probability of a UU continuation. Normalizing on the basis of any interview technology $j > 1$ would reduce the number of imputed UU continuations and thus increase the discrepancy between UU continuation probabilities and reported durations.[13] Using the first-interview definition of unemployment helps resolve the inconsistency between reported durations and observed UU continuations relative to a standardization based on any other interview technology.

The role of missing observations. Some have conjectured that rotation bias might arise from

[13] For example, the first-interview measure implies an unemployment-continuation probability of $\pi^*_{UU}/(1-\pi^*_{UM}) =$ 55.6% after correcting for rotation bias. By contrast, if we were to use rotation-bias-corrected transition probabilities $\pi^{**}_{X_1,X_2}$ in Table A-5 based on the average interview technology, we would calculate an implied unemployment-continuation probability of $\pi^{**}_{UU}/(1-\pi^{**}_{UM}) = 54.7\%$.

unemployed individuals exiting the sample more quickly than others. But rows 10-16 of Table 4 establish that the effect is instead directly related to the number of times the household has been interviewed. There are some people who were interviewed for the first time when the address would have been in rotation 2, for example because the individual moved into the address. The unemployment rate for these individuals is reported in row 11. Others were missing in both 1 and 2 and are being asked the questions for the first time in rotation 3 (row 12). For every group, we see the highest unemployment rate the first time people are asked the questions and a drop across each follow-up interview.[14]

Moreover, we observe an *increase* across rotations in the total number of individuals who are designated as not in the labor force (row 17 of Table 4). This cannot be people dropping out of the survey, but must come from some people changing their answers. Yet another way to get at this question is to look at the subset of individuals who gave answers in both rotation 1 and rotation 2. Row 1 of Table 5 shows that the unemployment rate for this group was 6.50% the first time they were asked the question and 6.23% the second time. Row 2 shows that among individuals who were sampled in both rotation 2 and rotation 3, the unemployment rate was 6.22% in rotation 2 and 5.99% in rotation 3.[15] The same pattern of the reported unemployment rate to drop among a fixed group of individuals whenever the household is asked the same questions a second time is seen in each of the subsequent rows of Table 5 as well.

Reconciliation with Krueger, Mas and Niu. The evidence in Krueger, Mas and Niu (2017) is sometimes interpreted as showing that rotation bias does not result from individuals being asked the same question multiple times. Krueger, Mas and Niu interpreted the duration of job search as a measure of the number of times an individual had previously reported being unemployed. But duration of job search is not a reliable indicator of the number of times people have answered the questions in earlier rotations. Of people in our sample who responded in both rotations 1 and 2, 30% of the U individuals in rotation 2 who reported unemployment durations 9 weeks or longer

[14] Indeed, the reported unemployment rate among people being asked the questions for the first time when in rotation 2 (7.4%) is even higher than the unemployment rate among people being asked the questions the first time when in rotation 1 (6.6%). This is a consequence of the fact that individuals who are M in some month of the survey are more likely than the general population to be U in the months when they are sampled.

[15] The average unemployment rate for the $j = 1$, u_{j+1} entry in Table 5 (6.23%) is not quite the same as the $j = 2$, u_j entry because the set of individuals who were not missing in either rotation 1 or 2 (which is the set of people who are counted in the $j = 1$ row) is not quite the same as the set of individuals who were not missing in either rotation 2 or 3 (which is the set of individuals who are counted in the $j = 2$ row).

had been counted as E or N in rotation 1 (4 weeks earlier). Krueger, Mas and Niu found that the biggest difference between rotations 1 and 2 comes from comparing people who report being unemployed with a duration less than 5 weeks ($U^{1.4}$) in rotation 1 with people who are $U^{1.4}$ in rotation 2. This is not an apples-to-apples comparison. In our 2001-2018 sample, the durations in rotation 1 are all solicited explicitly, whereas the durations for UU continuations into rotation 2 are imputed to be a number greater than 4 weeks. Thus by construction no one who is $U^{1.4}$ in rotation 2 could have been unemployed in rotation 1. Any statistic that conditions on not being U the previous month is selecting a subset of individuals who have a lower unemployment rate than the general population, which explains why $U^{1.4}$ in rotation 2 would be expected to be a smaller number than $U^{1.4}$ in rotation 1. Our data set contains a total of 39,000 individuals who were $U^{1.4}$ in rotation 1 but only 30,000 who were $U^{1.4}$ in rotation 2. By contrast, we have 28,000 $U^{5.14}$ in rotation 1 and 34,000 in rotation 2. This suggests that most of the "missing" $U^{1.4}$ in rotation 2 are being classified as $U^{5.14}$ on the basis of the BLS duration imputation but would have reported $U^{1.4}$ if allowed. The same pattern is seen in comparing rotations 5 and 6. Two-thirds of the drop in $U^{1.4}$ between 5 and 6 is accounted for by the rise in $U^{5.14}$.

Before 1994, durations for all individuals (including UU continuations) were directly solicited rather than imputed. A striking finding in Krueger, Mas and Niu's Figure 4 is that rotation bias among $U^{1.4}$ individuals was virtually nonexistent prior to 1994 and then appeared suddenly and dramatically when the BLS began imputing durations to UU continuations in 1994. Their figure shows that this break also coincides with a *decrease* in rotation bias in 1994 for $U^{5.14}$ individuals. We conclude that reported and imputed unemployment durations cannot be used in the way suggested by Krueger, Mas and Niu to identify the effects of being asked the survey questions multiple times.[16]

Conclusion. We again emphasize that the method we developed in equations (A17)-(A19) can be used to reconcile stocks with flows on the basis of any interview technology i. The evidence reviewed in this section leads us to recommend the first-interview concept of labor-force status

[16] Others have suggested that rotation bias might result from a difference between phone interviews and in-person interviews. For example, it is possible that respondents might want to impress the interviewer by showing their effort for job search when jobless, which would overstate the unemployment rate from personal interviews. However, the data suggest to us that this is an unlikely explanation. First, both the first and fifth rotation groups are typically surveyed in person, yet individuals in rotation 5 have significantly lower unemployment rates than those in rotation 1 (see column 6 of Table 1). Second, within rotation 5, individuals report significantly lower unemployment rates the more times they have previously been interviewed (see rows 11-14 of column 5 of Table 4). Third, rotation bias was observed during the time when all the interviews were conducted in person (see for example Hansen et al., 1955). For these reasons, we conclude that the mode of interview is unlikely to be the key explanation for rotation bias.

($i = 1$) as the one that should be used. The most important factor in rotation bias appears to be stigma and disengagement on the part of individuals who are unemployed or employed part time. Evidence in support of this hypothesis includes (1) characteristics of people claiming to be retired or disabled, (2) differences between self-reported answers and answers reported by proxy, (3) observed differences in unemployment rates and part-time employment rates across different demographic and education groups and different sample periods, and (4) comparing answers for missed interviews and matched interviews.

3.4 Estimates of labor-force participation and unemployment rates that correct for rotation bias.

Let $y_{X,t} = \sum_{j=1}^{8} y_{X,t}^{[j]}$ be the total number of individuals with labor-force status X in month t across all rotations j and π_t the (4×1) vector of resulting fractions $\pi_{X,t} = y_{X,t}/(y_{E,t} + y_{N,t} + y_{M,t} + y_{U,t})$. The unemployment rate for month t reported by BLS corresponds in our sample to $\pi_{U,t}/(\pi_{E,t}+\pi_{U,t})$. By contrast, the fractions for our preferred fixed-interview technology perspective are given by π_t^*, which would imply an unemployment rate of $\pi_{U,t}^*/(\pi_{E,t}^* + \pi_{U,t}^*)$. Our proposed adjustment to the unemployment rate at date t resulting from correcting for rotation bias is the difference between these two magnitudes:

$$\tilde{u}_t^R = \pi_{U,t}^*/(\pi_{E,t}^* + \pi_{U,t}^*) - \pi_{U,t}/(\pi_{E,t} + \pi_{U,t}). \tag{12}$$

This adjustment would increase the unemployment rate by 0.6% on average over our sample (see the first column of Table 6). The size of the adjustment for each month is indicated by the dotted blue line in the top panel of Figure 7.[17] This adjustment has a modest cyclical tendency, rising a little in the 2008-2009 and 2020 recessions as a result of the relation documented in the top panel of Figure 6. The adjustment also increases slightly over time as a result of the increasing trend in $\bar{\theta}_{NU,t}$ seen in Figure 5.

Correcting for rotation bias is even more important for purposes of calculating the labor-force

[17] Note that in order to preserve all the accounting identities, our raw data and all adjustments apply to seasonally unadjusted data. We have seasonally adjusted the corrections to display the values in Figure 7. The final adjusted estimates plotted in Figure 1 were calculated by adding the seasonally unadjusted correction like (12) to the seasonally unadjusted original BLS number and then seasonally adjusting the resulting sum. Seasonal adjustments in this paper were implemented using the X-11 procedure in RATS.

participation rate, as seen in the second column of Table 6 and the height of the dotted blue line in the second panel of Figure 7. This is because rotation bias arising from part-time employment is added to the rotation bias in unemployment in determining labor-force participation. The cyclical and trend components in the correction to the labor-force participation rate are slightly stronger than those for the unemployment rate.

The pandemic caused a huge increase in the number of missing individuals in 2020. We will discuss this phenomena in detail in the next section, but here comment briefly on its role in the correction for rotation bias. The challenges for completing interviews in the spring of 2020 were particularly serious for people being interviewed the first time compared to follow-up interviews which could be conducted by phone. Some readers might object that our fixed-interview normalization would be inappropriate for this episode insofar as the first-interview data may have become less reliable in the spring of 2020. This objection is based on a misunderstanding of our procedure. We use *all* the observations from *all* rotations observed in month t to construct our estimate π_t^* for that month. Our procedure in going from the observed data $\pi_t^{[j]}$ to the standardized measure π_t^* simply translates $\pi_t^{[j]}$ into first-interview-analogous answers based on the tendencies captured by $\bar{\theta}_t$. Because we construct $\bar{\theta}_t$ using strong smoothing, the correction is primarily determined by how rotation 1 answers differed from rotation j answers over the last several years prior to t with relatively little influence of the particular differences in period t. Our approach does not assume that rotation 1 answers in the spring of 2020 are "more reliable" than those for other rotations j. Instead, we are simply aggregating the answers given across all rotations j at t in terms of a unified coherent metric that has a fixed interpretation over time.

4 Missing individuals are different from the general population.

The summary statistics reported by the BLS for month t don't include estimates from individuals for whom no labor-force status was obtained in that particular month. This would be no problem if those missing from the survey are just like those included. However, the probabilities Π^* in Table 3 show that someone who is employed has a 6.6% probability of being missing in the next month, whereas someone who is unemployed has a 9.0% probability. Of those making ME, MN, or MU transitions, 5.9% are unemployed, although the unemployed only comprise 4.4% of the observed $E+N+U$ on average. In addition, of those making MU transitions, 60% claim that they have been searching for work longer than 4 weeks. In sum, missing individuals are more likely to be unemployed than a typical person in the sample.

Our M category includes the out-of-scope population, for example, people who leave the sample for reasons such as death, imprisonment, or enlistment in the army. Such individuals would show up in our data set as EM, NM, or UM transitions. We will not try to make any adjustment to labor-force measures for such individuals. Instead, our adjustments will be based solely on individuals who were M the previous month and are E, N, or U during the current month.[18]

To correct for the bias coming from nonrandom missing observations, we impute a labor-force status in month $t-1$ to individuals observed to make ME, MN, or MU transitions into period t. Our assumption is that some fraction $m_{E,t-1}$ of those missing in month $t-1$ are just like those who were counted as employed that month in terms of their transition probabilities, while fractions $m_{N,t-1}$ or $m_{U,t-1}$ share the same transition probabilities as those counted as N or U. We regard the remaining $m_{M,t-1} = 1 - m_{E,t-1} - m_{N,t-1} - m_{U,t-1}$ as "dormant observations" in the sense of having zero probability of being recorded as E, N, or U in month t.[19] The probabilities of observing ME, MN, and MU transitions would then be given by

$$\begin{bmatrix} \pi^*_{ME,t} \\ \pi^*_{MN,t} \\ \pi^*_{MU,t} \end{bmatrix} = \begin{bmatrix} \pi^*_{EE,t} & \pi^*_{NE,t} & \pi^*_{UE,t} \\ \pi^*_{EN,t} & \pi^*_{NN,t} & \pi^*_{UN,t} \\ \pi^*_{EU,t} & \pi^*_{NU,t} & \pi^*_{UU}, t \end{bmatrix} \begin{bmatrix} m_{E,t-1} \\ m_{N,t-1} \\ m_{U,t-1} \end{bmatrix}. \tag{13}$$

This system of equations can be solved to find $(m_{E,t-1}, m_{N,t-1}, m_{U,t-1})$. When solved for the full-sample average matrices π^* and Π^* in Table 3, the solutions are $(m_E, m_N, m_U) = (0.0996, 0.0489, 0.0123)$. Thus on average we impute 10% of missing observations to employed, 4.9% to not in the labor force, and 1.2% to unemployed. We make no imputation to the remaining 84% of those missing at $t-1$ because we did not get any new information (i.e., they were still missing) in t.

[18] This category includes individuals who were 15 in the previous month but became 16 in the current month, and those who were in the armed force in the previous month but now a civilian. However, we can directly observe these flows from the microdata, and the fractions of these observations are negligible (less than 0.1% of civilian non-institutional population). Hence, such individuals should not affect our estimates significantly.

[19] This would include people who are in the military, incarcerated, moved away from the address, or yet to move in, for example.

This adjustment is similar in spirit to that employed by Abowd and Zellner, with one critical difference. Their adjustments use observed transitions such as $\pi^{[j]}_{ME,t}$ that make no correction for rotation bias. By contrast, we use $\pi^{*}_{ME,t}$ which corrects for the fact that the status reported for rotation j would be expected to differ in systematic ways from the status in $j-1$.

Our assumption is that in normal times the characteristics of missing individuals do not change suddenly, so we exponentially smooth these as

$$\overline{m}_{X,t} = (1-\lambda)m_{X,t} + \lambda\overline{m}_{X,t-1} \quad \text{for } t = 2001{:}8 \text{ to } 2020{:}2$$

with the recursion started by setting $\bar{m}_{X,2001:7}$ equal to the average over the first year of the sample. We use $\lambda = 0.95$, which gives observations 1 year ago about half the weight as current observations for purposes of calculating $\overline{m}_{X,t}$.

The pandemic resulted in a huge increase in missing observations in the spring of 2020 for reasons that were very different from those that generate typical missing observations. To allow for the possibility of very rapid and dramatic changes in the characteristics of those missed, we do no smoothing for these data: $\bar{m}_{X,t} = m_{X,t}$ for $t = 2020{:}3$ to $2020{:}11$. We thus sacrifice some accuracy in the estimates for 2020 in order to gain flexibility in modeling the quick and dramatic changes in reasons why people may have been missed during this period. The resulting series for $\bar{m}_{X,t}$ are plotted in Figure 8. There are gradual increasing trends in $\bar{m}_{E,t}$ and $\bar{m}_{N,t}$ over the first twenty years followed by a dramatic but transitory spike in the spring of 2020. The countercyclical character of $\bar{m}_{U,t}$ is evident in both the Great Recession and the pandemic recession.

A useful way to summarize these is in terms of the fraction of unemployed among the M for whom a status of E, N, or U is imputed. This is given by the ratio $\bar{m}_{U,t-1}/(\bar{m}_{E,t-1} + \bar{m}_{N,t-1} + \bar{m}_{U,t-1})$ in the last panel of Figure 8. This highlights the countercyclical nature of the adjustment to the count of the unemployed as well as the fact that unemployed made up an unusually large share of the missing observations during the pandemic recession. In May of 2020, 4.46% of the individuals in rotations $j \in J$ were counted as employed in May but had been missing from $j-1$ in April.[20] This is 1.6 times the average fraction (2.77%) seen over 2001-2019, and is the primary

[20] That is, $\sum_{j\in J} y^{[j]}_{ME,2020:5} / \sum_{,i,k,j\in J} y^{[j]}_{ik,2020:5} = 0.0446$. Note that the actual calculation of m_{Et} is based on the full matrix of fixed-rotation parameters Π^{*}_{t} as described in equation (13). The numbers for rotations $j \in J$ are reported in the text in order to provide some understanding of the features of the raw data that lead m_{Et} to have

feature of the raw data that caused $\bar{m}_E$ in the first panel of Figure 8 to spike up. By contrast, the fraction of individuals who were M in April and U in May was 0.75%, three times the historically expected number. Because $\bar{m}_{MU,t}$ increased by a greater proportion than $\bar{m}_{EU,t}$, the ratio in the last panel of Figure 8 spikes up along with the levels in the first three.

Our estimates that correct for both rotation bias and nonrandomly missing observations are based on

$$\begin{bmatrix} \pi_{E,t}^{RM} \\ \pi_{N,t}^{RM} \\ \pi_{M,t}^{RM} \\ \pi_{U,t}^{RM} \end{bmatrix} = \begin{bmatrix} \pi_{E,t}^{*} + \bar{m}_{E,t}\pi_{M,t}^{*} \\ \pi_{N,t}^{*} + \bar{m}_{N,t}\pi_{M,t}^{*} \\ \pi_{M,t}^{*} - (\bar{m}_{E,t} + \bar{m}_{N,t} + \bar{m}_{U,t})\pi_{M,t}^{*} \\ \pi_{U,t}^{*} + \bar{m}_{U,t}\pi_{U,t}^{*} \end{bmatrix}.$$

Note that we only impute a labor-force status to those individuals who are M at t and either E, N, or U at $t+1$. If they are missing in both t and $t+1$, we make no imputation. Our combined adjustment to the unemployment rate in month t for rotation and missing bias is thus found from

$$\tilde{u}_t^{RM} = \pi_{U,t}^{RM}/(\pi_{E,t}^{RM} + \pi_{U,t}^{RM}) - \pi_{U,t}/(\pi_{E,t} + \pi_{U,t}).$$

On average over our sample, correcting for nonrandom missing observations adds an additional 0.3% to the unemployment rate and labor-force participation rate (see the third row in Table 6). It adds 0.8% to both unemployment and participation in August of 2020 (see the dashed lines in Figure 7).

5 Longer-term unemployed who are counted as not in the labor force.

To illustrate another striking inconsistency in the BLS data, consider the set of individuals who were counted as not in the labor force when they were interviewed in rotation 1 and counted as unemployed in rotation 2. In the second interview, respondents would be asked how long they have been looking for a job. The average percentage of these individuals who report they have been looking for work for τ weeks in this second month is indicated by the black bars in Figure 9.

---

the properties that it does.

Two-thirds of these people have a duration of unemployment in rotation 2 that is recorded to be longer than 4 weeks, despite the fact that the previous month they did not report actively looking for a job and so were counted as out of the labor force. 8.6% of NU individuals say that they have been looking for a job for 52 weeks and another 8.5% report having been looking for work for over 98 weeks. One is forced to conclude either that the job-search durations that people report do not reflect spells of true unemployment or that some of the individuals who are counted as not in the labor force should instead be counted as unemployed.

Forecasting evidence about NU^{5+} transitions. We first note that the number of people who will report a duration of unemployment in month t that is inconsistent with having been designated N at $t-1$ can be predicted quite accurately on the basis of data available at $t-1$. Let $N_{t-1}U_t^{5+}$ denote the total number of sampled individuals who were counted as not in the labor force in $t-1$ and unemployed with duration of 5 weeks or longer in month t. Let U_{t-1}^{5+} be the total number of individuals who were unemployed with reported duration of 5 weeks or longer in month $t-1$. Note that these numbers are obtained from two completely separate groups of people: U_{t-1}^{5+} comes from people who were unemployed in $t-1$, whereas $N_{t-1}U_t^{5+}$ comes from people who were not in the labor force in $t-1$. Note also that while the value of U_{t-1}^{5+} was known at $t-1$, the value of $N_{t-1}U_t^{5+}$ is not measured until t. Nevertheless, $N_{t-1}U_t^{5+}$ can be predicted quite accurately from U_{t-1}^{5+}. To summarize this predictability, we convert both numbers to population fractions. Let

$$m_{N,t-1}^{\sharp} = \frac{\sum_{\tau=5}^{99} \sum_{j \in J} y_{N,U,t}^{[j]}(\tau)}{\sum_{j \in J} \left[y_{E,t-1}^{[j-1]} + y_{N,t-1}^{[j-1]} + y_{M,t-1}^{[j-1]} + y_{U,t-1}^{[j-1]} \right]} \tag{14}$$

be the fraction of individuals who will make an NU^{5+} transition between $t-1$ and t and let $u_{t-1}^{5+} = U_{t-1}^{5+}/(E_{t-1} + U_{t-1})$ be the medium- to long-term unemployment rate reported by the BLS in month $t-1$. Expressing these fractions as percentages, consider the following OLS regression (Newey-West standard errors with 12 lags in parentheses):

$$m_{N,t-1}^{\sharp} = \underset{(0.0018)}{0.0690} + \underset{(0.0079)}{0.0682} u_{t-1}^{5+} + \hat{\varepsilon}_{t-1} \qquad R^2 = 0.91. \tag{15}$$

Whenever one observes a high value of u_{t-1}^{5+} one can quite accurately predict that many of the individuals who are classified as N_{t-1} that month will report in the following month that they have been looking for work for longer than 5 weeks. This observation invites us to ask whether some of the individuals counted as N_{t-1} should have been included with the group of unemployed at date $t-1$.

We can also examine at the individual level what a person's report of $N_{t-1}U_t^{5+}$ predicts about their own labor-force status at $t+1$. The first column of Table 7 examines UUU continuations in months $t-2$, $t-1$, and t for which the reported durations would be consistent with a true UUU continuation.[21] As we go down the rows, the history is consistent with a longer initial duration in month $t-2$. Our framework would predict that the probability of being employed in month $t+1$ would decrease as we move down the rows. This is because individuals who have a lower ex ante probability of becoming employed at $t+1$ make up a larger fraction of the pool at t as we move down the rows.[22] This is exactly what we observe in the data. The third column looks at individuals with an intervening N status in month $t-1$ but with the same U durations in $t-2$ and t as in column 1. These probabilities also tend to decrease as we move down rows. The job-finding prospects for someone who begins a UUU stretch with reported initial duration of 5 to 14 weeks (16%) is similar to that for somebody who begins a UNU stretch with duration 5-14 weeks (15%), as are the probabilities for someone beginning with more than 26 weeks (8% versus 7%, respectively). This again is consistent with the view that the intervening N should instead be viewed as part of a spell of continued unemployment.

Demographic evidence on NU^{5+} transitions. It is also interesting to examine how $m_{Nt}^{\sharp}$ varies across different demographic groups. We summarize this evidence in the top panel of Figure 10. Over the 2001:7-2008:6 subsample, $m_{Nt}^{\sharp}$ averaged 0.44% for those without a high school diploma but only 0.15% for college graduates, a fact that could lead some to conclude that better educated individuals are less prone to misreport their time spent searching for a job. But we also observe that the value of $m_{Nt}^{\sharp}$ is higher for every group in the 2008:7-2014:6 subsample, correlating very strongly with the higher rates of u_t^{5+} for every group in this subsample. For example, for college graduates the average values of u_t^{5+} and $m_{Nt}^{\sharp}$ during the Great Recession were about the same as the average values of these two variables for the overall population prior to the recession. When their unemployment rate is high, college graduates behave just as the rest of the population does

[21] For example, $U_{t-2}^{1.4}, U_{t-1}^{5.14}, U_t^{5.14}$ refers to someone who reported being newly unemployed in $t-2$ and being unemployed between 5 and 14 weeks in $t-1$ and t.

[22] See Ahn and Hamilton (2020).

in normal times. Both $u^{5+}_{cg,t}$ and $m^{\sharp}_{cg,Nt}$ returned to their historical values when the recession was over. A regression fit to the 27 data points in the top panel of Figure 10 is

$$m^{\sharp}_{i,NT} = \underset{(0.0251)}{0.0629} + \underset{(0.0045)}{0.0609} u^{5+}_{iT} + \hat{\varepsilon}_{iT} \qquad R^2 = 0.88. \tag{16}$$

Almost all of the variation in observed NU^{5+} transitions across demographic groups and across time can be explained by the simple hypothesis that when more people are experiencing longer-term unemployment, the number of unemployed who are incorrectly classified as not looking for a job is higher.

Note moreover that the coefficients from the demographic regression (16) are remarkably similar to those for the aggregate time-series forecasting relation (15). We illustrate this point visually in the bottom panel of Figure 10. For each month $t-1$, we would predict on the basis of the demographic regression a value for next month's aggregate $N_{t-1}U^{5+}_t$ transitions to be $0.0629 + (0.0609)u^{5+}_{i,t-1}$. The figure compares this with the actual transitions $m^{\sharp}_{N,t-1}$. The predicted $\hat{m}^{\sharp}_{N,t-1}$ is on average lower than the actual $m^{\sharp}_{N,t-1}$, as a result of the fact that A1-A3 fall a little above the demographic regression line in the top panel. Notwithstanding, the fit is quite remarkable. One can predict $N_{t-1}U^{5+}_t$ transitions quite accurately, for any demographic group or any point in time, solely on the basis of u^{5+}_{t-1}.

The coefficients for this forecasting relation were estimated from cross-section data prior to the 2020 recession, and we circulated an earlier version of this paper that proposed using the cross-sectional coefficients in this way. Subsequent to circulation of our paper, the U.S. experienced a pandemic in 2020 that brought about striking changes in unemployment and BLS measurement. This additional year of observations provide this relation with an extremely challenging true out-of-sample test. As seen in the bottom panel of Figure 10, the coefficients estimated from the cross-sectional relation prior to 2020:1 do quite an amazing job at predicting the wild changes in 2020:2-2020:12. We view this as further evidence supporting the view that $N_{t-1}U^{5+}_t$ individuals should be placed in the same category as those officially counted as unemployed in $t-1$.

Additional evidence on NU^{5+} transitions.[23] There are a number of other factors that reinforce this conclusion. The first factor is the individual's own retrospective assessment. When asked at

[23] The numbers in this paragraph are based on an analysis of 2001:7 to 2020:2 data.

time t, "how long have you been looking for work?", if the individual answers more than 4 weeks it indicates that the individual's own perception at t is fully consistent with characterizing them as U at $t-1$. Second, the objective probability of being employed the next period is similar across the two groups: $P(E_{t+1}|N_{t-1}, U_t^{5+}) = 12.5\%$ versus $P(E_{t+1}|U_{t-1}, U_t^{5+}) = 15.5\%$, in sharp contrast for example to $P(E_{t+1}|E_{t-1}, U_t^{5+}) = 37.8\%$. Third, information the individuals gave at $t-1$ would also identify many of the $N_{t-1}U_t^{5+}$ transitions as more attached to the labor force than typical N_{t-1}. Specifically, people who are not in the labor force are asked whether they want a job. Only 5.3% of all N_{t-1} answered this question yes, whereas 44% of $N_{t-1}U_t^{5+}$ answered the question yes at $t-1$. The indication that a person wants a job (WJ) is furthermore an objective predictor that they will find one. For example, $P(E_{t+1}|N_{t-1}^{WJ}, U_t) = 14.1\%$ versus $P(E_{t+1}|U_{t-1}, U_t) = 15.5\%$.

Based on these considerations, our recommendation is to classify observed $N_{t-1}U_t^{5+}$ transitions as having been U rather than N at $t-1$. This adjustment is closely related to that recommended by Rothstein (2011), Elsby et al. (2011), Elsby, Hobijn, and Şahin (2015), and Farber and Valletta (2015) who reclassified all UNU as UUU. Our approach differs from theirs in that we utilize the reported duration of unemployment when correcting classification errors. The adjustment just described would only classify UNU as UUU if the final U reports a duration of job search greater than 4 weeks.

Implications of $N_{t-1}U_t^{5+}$ transitions for unemployment and labor-force participation. The number of individuals we have been discussing in this section is very large; the value of $m_{N,t-1}^{\sharp}$ in (14) averages 0.37% of all individuals over the full sample and reaches a maximum of 0.68% in April 2020. Note moreover that we have normalized $m_{N,t-1}^{\sharp}$ as a fraction of all $E+N+M+U$. To calculate the approximate effect on the reported unemployment rate of counting $N_{t-1}U_t^{5+}$ as unemployed at $t-1$, we would multiply $m_{N,t-1}^{\sharp}$ by 2.2.[24] Our proposed adjustment thus would increase the unemployment rate by about $(2.2)(0.37) = 0.8\%$ on average and by $(2.2)(0.68) = 1.5\%$ in April 2020.

Before considering the full details of these adjustments we want to explore some of the additional

[24] That is,

$$\frac{\sum_{j\in J}\left[y_{E,t-1}^{[j-1]} + y_{N,t-1}^{[j-1]} + y_{M,t-1}^{[j-1]} + y_{U,t-1}^{[j-1]}\right]}{\sum_{j\in J}\left[y_{E,t-1}^{[j-1]} + y_{U,t-1}^{[j-1]}\right]}$$

averages 2.2 over the sample. For details of the actual adjustment to the unemployment rate each month, see Section 6.6.

implications of this finding. One issue it raises is that some of the observed UN transitions are actually continuing spells of unemployment. Moreover, it is not clear how we should interpret the durations of job search reported by the unemployed and how those self-reports are related to the true objective durations of measured spells of unemployment. We develop a complete reconciliation of stocks, flows, and reported durations in the next section.

6 Reconciliation of stocks and flows with reported durations.

We begin with a statistical description of the job-search durations reported by the unemployed.

6.1 Characteristics of reported durations.

We first examine the durations of unemployment that are reported on average over our sample by people who are being interviewed for the first time (rotation 1). The black bars in the top panel of Figure 13 show the percentage of individuals in rotation 1 who report being unemployed with the indicated duration of job search in weeks. Clearly there are some significant reporting errors arising from number preference. Respondents are more likely to report spells as an integer number of months, and for longer spells as either 6 months, 1 year, 18 months, or longer than 99 weeks. For shorter spells, people are more likely to report an even number of weeks instead of an odd number; for example, on average there are more people reporting 2 weeks than 1 and 6 weeks than 5. Respondents are extremely unlikely to report a duration of zero weeks, and for this reason we group the 0-week and 1-week observations together into a category of reported duration less than or equal to one week.

To interpret these numbers in an internally consistent way, we impose the restriction that the only way an individual could have been unemployed for τ weeks would be if the individual had been unemployed for $\tau - 1$ weeks the week before. Thus if $\pi_U^\dagger(\tau)$ denotes an internally consistent summary of the fraction of the population who have been searching for τ weeks, the function $\pi_U^\dagger(\tau)$ should be monotonically decreasing in τ. There are a number of monotonically decreasing parametric functions one could use. We have found that the best fit to the data comes from using a mixture of two exponential distributions with mixing portions w_1 and $(1 - w_1)$ and decay rates p_1 and p_2, respectively. As we will see below, this functional form is also convenient because it allows us to make some easy comparisons between reported duration of job search and objective unemployment continuation probabilities. We form a (99×1) vector $\pi_U^\dagger$ whose τth element for $\tau = 1, 2, ..., 98$ is an internally consistent representation of the fraction of the working-age population who perceive having been unemployed for a duration of τ weeks at a fixed point in time, while the 99th element is the fraction with perceived duration greater than 98 weeks:

$$\pi_U^\dagger = \pi_{1U}^\dagger + \pi_{2U}^\dagger \tag{17}$$

$$\underset{(99\times 1)}{\pi_{iU}^\dagger} = \pi_U w_i (1 - p_i) \begin{bmatrix} 1 & p_i & p_i^2 & \cdots & p_i^{97} & p_i^{98}/(1-p_i) \end{bmatrix}' \quad \text{for } i = 1, 2. \tag{18}$$

Here π_U denotes the fraction of the population who are unemployed and w_i the fraction of those individuals who are type i. Such a distribution would be the outcome of a steady state in which there was a fraction $\pi_U w_1(1 - p_1)$ of the population who lose their jobs each week and for each of whom the probability of continuing unemployed in any subsequent week is p_1, and an additional inflow of $\pi_U w_2(1 - p_2)$ individuals with continuation probability p_2.[25]

We allow for the various forms of number preference noted above by introducing a (99×99) matrix $A(\theta_A)$ whose elements are determined by a (13×1) vector θ_A. The first element $\theta_{A,1}$ allows a preference for reporting short durations as an even rather than an odd number of weeks, assuming that someone whose true duration is $\tau = 1, 3, 5$, or 7 in fact reports duration 2, 4, 6, or 8 with probability $\theta_{A,1}$. The value of $\theta_{A,2}$ represents the probability that someone will round their duration up or down by a week to reach an integer number of months for durations within one week of 1, 2, 3 or 4 months, while someone two weeks away from either of two months is presumed to round down with probability $\theta_{A,3}/2$ and up with probability $\theta_{A,3}/2$. As we move to longer durations we allow for the possibility that the rounding tendencies become stronger, introducing new pairs of parameters for durations between 5-7 months, 8-11 months, or 12 or more months. The last elements of θ_A allow for preferences for integer multiples of 6 months for longer durations. For each τ the τth column of A sums to unity and characterizes the probability that someone whose true duration category is τ will report each of the possible categories i between 1 and 99, where i

[25] We will later examine some testable implications of such an interpretation by looking at the actual unemployment-continuation probabilities for different individuals and also look at alternative functional forms. But for now we propose (17) and (18) as a simple but flexible parametric functional form with which to impose monotonicity on $\pi_U^\dagger(\tau)$.

or $\tau = 99$ is interpreted as true or reported durations longer than 98 weeks. Appendix B provides more details on the structure we use to represent the matrix A. Note that our framework does not impose the assumption of the existence or magnitude of any particular reporting error, as it includes as a special case no reporting error of any kind when $\theta_A = 0$.

Let $y_{X,t}^{[1]}$ be the number of individuals in rotation 1 sampled at date t who report status X for X one of E (employed), N (not in labor force), M (labor-force status for that individual is missing), or U (unemployed). We summarize further detail in the last category in terms of $y_{U,t}^{[1]}(\tau)$ which is the number of unemployed who report having been looking for work for τ weeks for $\tau = 1, ..., 99$.[26] We compare the observed values $y_{U,t}^{[1]}(\tau)$ with the predicted values represented by the (99×1) vector

$$\dot{\pi}_U = A\pi_U^{\dagger}. \tag{19}$$

We also let π_X denote the overall fraction of the population reporting status $X \in \{E, N, M, U\}$. If we treated observations as independent across months t the log likelihood of the rotation 1 observations alone would then be

$$\begin{aligned} \ell_X^{[1]}(\theta_X) &= \textstyle\sum_{t=1}^T [y_{E,t}^{[1]} \ln \pi_E + y_{N,t}^{[1]} \ln \pi_N + y_{M,t}^{[1]} \ln \pi_M] \\ &\quad + \textstyle\sum_{t=1}^T \sum_{\tau=1}^{99} y_{U,t}^{[1]}(\tau) \ln \dot{\pi}_U(\tau). \end{aligned} \tag{20}$$

We can maximize this with respect to $\theta_A, p_1, p_2, w_1, w_2, \pi_E, \pi_N, \pi_M, \pi_U$ subject to the constraint that all probabilities are between 0 and 1 and sum to unity.[27]

Estimates are reported in column 1 of Table 8, along with quasi-maximum-likelihood standard errors in column 2 which allow for the possibility that $y_{X,t}^{[1]}$ is correlated across time (calculated as described in Appendix C). The predicted values $\dot{\pi}_U(\tau)$ are compared with the average reported values in the top panel of Figure 11.[28] This framework is able to describe the reported values extremely accurately. The estimated latent function $\pi_U^{\dagger}(\tau)$ along with its two contributing com-

[26] The duration is top-coded at 99 weeks in our data.

[27] Maximum likelihood estimates of some parameters are known analytically. Let $y_X = \sum_{t=1}^T y_{X,t}$ denote the total number of observations in category X and $n = (y_E + y_N + y_M + y_U)$ the total number of observations. Then $\hat{\pi}_X = y_X/n$ for $X \in \{E, N, M, U\}$. These values can be substituted into expression (20) and the resulting concentrated likelihood then maximized with respect to θ_A, p_1, p_2, w_1 with $w_2 = 1 - w_1$.

[28] As noted in the previous footnote, by the nature of the maximization problem, the estimated values $\hat{\pi}_X$ for $X = E, N, M$ exactly match the historical fractions $y_X/(y_E + y_N + y_M + y_U)$.

ponents are plotted as a function of τ in the bottom panel of Figure 11. We also considered an alternative functional form based on a Weibull distribution, as discussed in Appendix D. The mixture of exponentials has a much better fit to the data than that for the Weibull specification, and we will use it in our baseline analysis.

For rotations 2-4 and 6-8, BLS imputes a duration to those reporting UU continuations, making durations for these individuals a hybrid of perceived and imputed quantities. This can create a downward bias in the number of individuals unemployed for less than 5 weeks as discussed by Abraham and Shimer (2001) and Shimer (2012), and blurs the inconsistency between perceived and imputed durations. Since our goal is to characterize perceived durations separately from objective durations, we do not use the imputed duration in the second month in unemployment. However, there are no imputations for unemployment durations for those people in rotation 5. We therefore repeated the analysis with $y_{X,t}^{[1]}$ in (20) replaced by $y_{X,t}^{[5]}$. Parameter estimates and standard errors are reported in columns 3 and 4 of Table 8. These are very similar to those inferred from the rotation 1 observations alone.

6.2 Characteristics of NU, EU, and MU transitions.

We now return to the set of people in Figure 9 who reported that they have been looking for work for more than 4 weeks despite having been designated as N the previous month. Our question is, What distribution characterizes their reported duration of job search? We represent the probability of transitions from N to E, N, M , or U with parameters $\pi_{NE}, \pi_{NN}, \pi_{NM}, \pi_{NU}$, respectively, where these four numbers sum to unity. Of those who make an NU transition and report an unemployment duration greater than 4 weeks, suppose that their perceived duration can again be represented by a mixture of two exponentials with decay parameters $p_{1,NU}$ or $p_{2,NU}$. We assume that some fractions $q_{1,NU}, q_{2,NU}, q_{3,NU}$, and $q_{4,NU}$ of those making the NU transition will perceive their unemployment duration to be 1,2,3, or 4 weeks respectively, treating these values of $q_{j,NU}$ completely unrestrained. A fraction $q_{5,NU}$ perceive a duration greater than 4 weeks drawn from an exponential distribution with parameter $p_{1,NU}$ and a fraction $q_{6,NU}$ are characterized by

$p_{2,NU}$, with $\sum_{j=1}^{6} q_{j,NU} = 1$. We thus calculate

$$\pi^{\dagger}_{NU}(\tau) = \begin{cases} q_{\tau,NU} & \text{for } \tau = 1,2,3,4 \\ q_{5,NU}(1-p_{1,NU})p_{1,NU}^{\tau-5} + q_{6,NU}(1-p_{2,NU})p_{2,NU}^{\tau-5} & \text{for } \tau = 5,6,...,98 \\ q_{5,NU}p_{1,NU}^{94} + q_{6,NU}p_{2,NU}^{94} & \text{for } \tau = 99 \end{cases}. \quad (21)$$

The predicted probability of each reported duration is then given by $\dot{\pi}_{NU} = \pi_{NU} A \pi^{\dagger}_{NU}$.

Let $y^{[2]}_{NX,t}$ denote the number of individuals who were counted as not in the labor force in rotation 1 in month $t-1$ and reported status X at date t where $X \in \{E, N, M, U\}$. Let $y^{[2]}_{NU,t}(\tau)$ denote the number of NU who report unemployment duration $\tau \in \{1, ..., 98, \geq 99\}$ in rotation 2. Then the contribution to the likelihood for months $t = 1, ..., T$ from rotation 2 NX transitions is

$$\begin{aligned} \ell^{[2]}_{NX}(\theta_{NX}) &= \textstyle\sum_{t=1}^{T} [y^{[2]}_{NE,t} \ln \pi_{NE} + y^{[2]}_{NN,t} \ln \pi_{NN} + y^{[2]}_{NM,t} \ln \pi_{NM}] \\ &\quad + \textstyle\sum_{t=1}^{T} \sum_{\tau=1}^{99} y^{[2]}_{NU,t}(\tau) \ln \dot{\pi}_{NU}(\tau). \end{aligned} \quad (22)$$

This expression can then be maximized with respect to $\theta_{NX} = (\theta'_{A,NU}, p_{1,NU}, p_{2,NU}, \pi_{NE}, \pi_{NN}, \pi_{NM}, \pi_{NU}, q_{1,NU}, q_{2,NU}, ..., q_{6,NU})'$ subject to the constraints that all parameters fall between 0 and 1, $\pi_{NE} + \pi_{NN} + \pi_{NM} + \pi_{NU} = 1$ and $\sum_{j=1}^{6} q_{j,NU} = 1$.

Quasi-maximum-likelihood estimates $\hat{\theta}_{NX}$ are reported in column 5 of Table 8 and predicted values $\dot{\pi}_{NU}$ compared with historical average values for y_{NU} in Figure 9. Note that Figures 9 and 11 come from completely different people. The individuals in Figure 11 were reported to be unemployed when they were surveyed in rotation 1 whereas the individuals in Figure 9 were reported to be N when they were surveyed in rotation 1. Despite the difference in people and status, the similarity in the figures is striking, and the parameters in columns 1 and 5 of Table 8 are very similar. In particular, p_2, the key parameter summarizing perceived duration for the long-term unemployed, is 0.9740 for $N_{t-1}U_t^{5+}$ individuals and 0.9732 for U_{t-1}. This similarity offers additional support for our recommendation to count $N_{t-1}U_t^{5+}$ individuals as having been unemployed in $t-1$.

Next consider the status in month t of individuals who were recorded as employed when sampled in rotation 1 in month $t-1$. Twenty-six percent of those who make EU transitions report durations longer than 4 weeks. Unlike the NU transitions, we do not interpret these as necessarily implying an inaccuracy in either the E or U designation. Kudlyak and Lange (2018) noted these could represent records of individuals who were employed in $t-1$ but were engaged in on–the-job search for a new job.[29] But this observation certainly establishes that we cannot regard reported job-search durations by themselves as a valid indicator of the length of an uninterrupted spell of unemployment. When we maximize (22) with NX replaced by EX we arrive at the estimates reported in columns 7 and 8 of Table 8. Far fewer EU transitions describe themselves as long-time job seekers ($q_{6,EU} = 0.18$ versus $q_{6,NU} = 0.51$).

We also looked at the status in rotation 2 of individuals who were missing in rotation 1, replacing EX with MX. Quasi-maximum-likelihood estimates are reported in column 9 of Table 8. Individuals making MU transitions look similar to those making NU transitions and the pool of unemployed in rotation 1.

6.3 Characteristics of UX transitions.

We next examine UX transitions. The bars in the top panel of Figure 12 show $\check{\pi}_U(\tau)$, the observed probability that someone in rotation 1 who reports being unemployed with duration τ weeks will still be unemployed the following month.[30] This probability rises as a function of duration before eventually plateauing at a value around 0.62 for durations over half a year. One way this feature of the data is often captured is by defining some arbitrary cutoff K with any duration $\tau \leq K$ designated as short-term unemployed who are assumed to have some continuation probability $\gamma_{1,UU}$ while long-term unemployed ($\tau > K$) have a different probability $\gamma_{2,UU}$. That kind of simple dichotomization into short-term and long-term unemployment would have the drawbacks that it requires picking an arbitrary cut-off K and implies an abrupt discontinuity in outcomes expected for individuals slightly below K relative to those slightly above K.

Our parameterization suggests a smooth function that could be used as a natural alternative

[29] Elsby et al. (2011) and Farber and Valletta (2015) suggested that EU^{5+} individuals could also be reporting the time since the last salient job. Both this interpretation, as well as that of Kudlyak and Lange (2018), support the conclusion that the reported duration associated with an EU transition should not be interpreted as the duration of an uninterrupted spell of unemployment. Ahn and Shao (2021) further documented that on-the-job search constitutes a non-negligible fraction of aggregate job search. Hall and Kudlyak (2019) found that many job losers make frequent transitions between short-term employment, unemployment, and out of the labor force before finding a long-term job.

[30] To avoid plotting values for observations with excessive sampling error, we set this probability to 0 for durations with 10 or fewer observations over the whole sample.

to an arbitrary cutoff. We have summarized the distribution of reported durations for those unemployed in rotation 1 as coming from a mixture of two types of individuals, where type 1 report numbers consistent with a weekly continuation probability of p_1 and type 2 have a reported continuation probability of p_2. We modeled the fraction of the population that reports being unemployed with duration τ as given by the τth element of the vector $\xi_1 + \xi_2$ where $\xi_i = A\pi^{\dagger}_{iU}$ for $\pi^{\dagger}_{iU}$ given in (18). If we observe someone reports a duration of τ, the probability that the individual is type i is obtained from the formula

$$\eta_i(\tau) = \xi_i(\tau)/[\xi_1(\tau) + \xi_2(\tau)] \tag{23}$$

for $i = 1$ or 2. The function $\eta_2(\tau)$ is plotted in the bottom panel of Figure 12.[31] Someone who reports a duration of $\tau = 1$ week is quite unlikely to have come from the second distribution, whereas someone who reports a duration greater than 40 weeks is almost certain to have come from the second distribution.[32]

This formula allows us to estimate the actual observed monthly transition probabilities for the two types. Let $\gamma_{i,UX}$ be the probability that an individual of type i makes a transition from unemployment in rotation 1 to status $X = E, N, M,$ or U in rotation 2, so $\gamma_{i,UE} + \gamma_{i,UN} + \gamma_{i,UM} + \gamma_{i,UU} = 1$ for both $i = 1$ and $i = 2$. Let η_i denote the vector whose τth element is $\eta_i(\tau)$ and $\dot{\pi}_{UX}$ the (99×1) vector whose τth element is the observed probability that someone who reports duration τ in month t has status X in month $t+1$. Under the above assumptions $\dot{\pi}_{UX}$ would be predicted to be

$$\dot{\pi}_{UX} = \eta_1 \gamma_{1,UX} + \eta_2 \gamma_{2,UX}. \tag{24}$$

Let $y^{[2]}_{UX,t}(\tau)$ denote the observed number of individuals who report U with duration τ in rotation 1 and status X in rotation 2. We then have the likelihood function

$$\begin{aligned} \ell^{[2]}_{UX}(\theta_{UX}) = \textstyle\sum_{t=1}^{T} \sum_{\tau=1}^{99} \quad & [y^{[2]}_{UE,t}(\tau) \ln \dot{\pi}_{UE}(\tau) + y^{[2]}_{UN,t}(\tau) \ln \dot{\pi}_{UN}(\tau) \\ & + y^{[2]}_{UM,t}(\tau) \ln \dot{\pi}_{UM}(\tau) + y^{[2]}_{UU,t}(\tau) \ln \dot{\pi}_{UU}(\tau)]. \end{aligned} \tag{25}$$

[31] For purposes of this graph, this function was calculated using the values of w_1, p_1, p_2, θ_A from Table 10, which pool all observations from all rotations to estimate these parameters.

[32] The function dips down at duration $\tau = 26$ weeks because, given the tendency of answers to clump at this value, this observation includes many individuals whose true duration is less than 26 weeks and accordingly contains a higher mix of type 1 relative to those reporting 25 weeks.

We fixed η_2 to be the function plotted in the bottom panel of Figure 12 and maximized (25) with respect to $\{\gamma_{i,UE}, \gamma_{i,UN}, \gamma_{i,UM}, \gamma_{i,UU}\}_{i=1,2}$ subject to the constraint that $\gamma_{i,UE} + \gamma_{i,UN} + \gamma_{i,UM} + \gamma_{i,UU} = 1$ for $i = 1, 2$.

Quasi-maximum-likelihood estimates and standard errors are reported in rows 1 and 2 of Table 9. Type 1 individuals have a 32% probability of being employed next month, whereas the probability for type 2 individuals is only 12%. Type 1 individuals have a 37% probability of being unemployed next month, whereas for type 2 the probability is 57%. The red line in the top panel of Figure 12 show the predicted values for the unemployment-continuation probability implied by these maximum likelihood estimates.[33] This function provides a very good summary of the raw data.

We also repeated the analysis using only data for individuals who were unemployed in rotation 5, with very similar results shown in rows 3 and 4. Our preferred estimates pool together all observations for all rotations but still estimate $\gamma_{i,UU}$ completely independently of the value of p_i, while treating the values of θ_A, p_1, and p_2 as the same across all rotation groups. This summary of the full data set was obtained by maximizing the full-sample likelihood

$$\ell = \ell_X^{[1]} + \ell_X^{[5]} + \textstyle\sum_{j \in J} \left(\ell_{EX}^{[j]} + \ell_{NX}^{[j]} + \ell_{MX}^{[j]} \right) + \ell_{UX}^{[2]} + \ell_{UX}^{[6]}. \tag{26}$$

These full-sample estimates are reported in Table 10.

Now let us compare the estimated objective unemployment-continuation probability for type 1 individuals ($\gamma_{1,UU}$) with the value that would be predicted on the basis of their reported durations. If type 1 individuals truly had a weekly unemployment-continuation probability of $p_1 = 0.8115$, we would expect to observe a monthly continuation probability of $0.8115^{4.33} = 0.40$. If we condition on missing observations having the same distribution as observed E, N and U, this value turns out to equal exactly the value we'd predict from Table 10 of $\gamma_{1,UU}/(1 - \gamma_{1,UM}) = 0.40$. Note that our approach did not impose this in any way; $\hat{p}_1$ is based solely on reported durations, whereas $\hat{\gamma}_{1,UU}$ is based solely on observed continuations. The exercise shows that the durations reported by type 1 individuals are entirely consistent with the observed labor-force flows for those individuals.

By contrast, the long-term unemployed are another story. Their perceived weekly

[33] That is, the red line plots $\eta_1(\tau)\hat{\gamma}_{1,UU}/(1 - \hat{\gamma}_{1,UM}) + \eta_2(\tau)\hat{\gamma}_{2,UU}/(1 - \hat{\gamma}_{2,UM})$ as a function of τ.

unemployment-continuation probability of $p_2 = 0.9727$ would imply a monthly continuation probability of $0.9727^{4.33} = 0.89$, far larger than the estimate $\gamma_{2,UU}/(1-\gamma_{2,UM}) = 0.62$. Even more dramatically, a monthly continuation probability of 0.62 would mean a probability of remaining unemployed for 6 months of $0.62^6 = 0.06$. But in the BLS data, the fraction of those unemployed who report durations over 26 weeks averages 28%. Far fewer people than are reported in the data should be unemployed longer than 6 months if people left the pool of long-term unemployed at anything like the rate implied by $\gamma_{2,UU}$. The observed unemployment continuation probabilities are not consistent with the distribution of reported unemployment durations.

That conclusion is robust whether one uses our parametric model or any other. For example, Appendix D derives the analogous result using a Weibull characterization of durations. Any model that accurately describes the cross-section of durations – and ours does so quite well – will predict an unemployment-continuation similar to the stock-based measure plotted as the black line in Figure 1, which we noted is inconsistent with flow-based measures in green. The main advantage of our parametric approach is that it highlights that this inconsistency between the stock-based and flow-based measures comes entirely from those whom we have characterized as the perceived long-term unemployed. This insight is new to this literature.

6.4 Changes in parameters over time.

So far in this section we have calculated parameter estimates by maximizing the likelihood of the full sample of observations. We now generalize this to allow these parameter estimates to change over time. Trying to estimate all 37 parameters separately for each month does not result in very reliable estimates, so we again adapt the idea of exponential smoothing. Let $\bar{y}^{[j]}_{X,t}$ denote the smoothed status counts in equation (9). and $\bar{y}^{[j]}_{X_1,X_2,t}(\tau) = \lambda\bar{y}^{[j]}_{X_1,X_2,t-1}(\tau)+(1-\lambda)y^{[j]}_{X_1,X_2,t-1}(\tau)$ corresponding smoothed transition counts. For each t we maximize a one-period likelihood function of the form of (26)

$$\bar{\ell}_t = \bar{\ell}^{[1]}_{X,t} + \bar{\ell}^{[5]}_{X,t} + \sum_{j\in J}\left(\bar{\ell}^{[j]}_{EX,t} + \bar{\ell}^{[j]}_{NX,t} + \bar{\ell}^{[j]}_{MX,t}\right) + \bar{\ell}^{[2]}_{UX,t} + \bar{\ell}^{[6]}_{UX,t}$$

where for example

$$\bar{\ell}_{X,t}^{[1]} = \bar{y}_{E,t}^{[1]} \ln \pi_{E,t} + \bar{y}_{N,t}^{[1]} \ln \pi_{N,t} + \bar{y}_{M,t}^{[1]} \ln \pi_{M,t} + \sum_{\tau=1}^{99} \bar{y}_{U,t}^{[1]}(\tau) \ln \dot{\pi}_{U,t}(\tau).$$

If the smoothing parameter $\lambda = 1$, the estimate for $t = T$ corresponds to the full-sample MLE reported in Table 10, whereas if $\lambda = 0$, the parameter estimates would correspond to maximizing the likelihood function separately for each month. We allow these parameters to change more quickly than those characterizing rotation bias, setting $\lambda = 0.95$. Exponential smoothing here is similar to basing estimates on a rolling sample of the most recent two years of observations as of date t, except that unlike a simple rolling window, observations one year ago count only half as much as data for t and observations two years ago about a quarter as much. Also unlike a rolling window, observations more than two years previous continue to be given some small weight.

The estimates for a few key parameters are shown in Figure 13. The parameter $\bar{p}_{1,t}$, which summarizes the reported unemployment durations of type 1 workers, exhibits significant seasonality. It fell sharply in April 2020 in response to the surge of individuals who reported being unemployed for just a few weeks, rose as many of those individuals remained unemployed over the next few months, and fell back down as the number of newly unemployed declined during the recovery. There is also a less dramatic decline in $\bar{p}_{1t}$ during the recovery from the Great Recession. By contrast, $\bar{p}_{2t}$, which reflects reported durations of the longer-term unemployed, continued to rise in the slow recovery from the Great Recession but fell in the rapid recovery from the pandemic recession of 2020. The value of $\bar{p}_{2t}$ also exhibits an increasing trend over time. The share of total unemployment coming from the long-term unemployed $\bar{w}_{2t}$ rose in the slow recovery from the Great Recession and fell in the initial months of the pandemic recession. This behavior is paralleled in $\bar{q}_{6,NU,t}$ and $\bar{q}_{6,MU,t}$ which capture the share of reported long-term unemployed among NU and MU transitions.

Note that the estimates $\gamma_{i,UX,t}$ are based on results from pooling transitions from rotation 1 to 2 with those from rotation 5 to 6. Letting $\gamma_{i,t} = (\gamma_{i,UE,t}, \gamma_{i,UN,t}, \gamma_{i,UM,t}, \gamma_{i,UU,t})'$, we can convert these to interview-one transition probabilities using

$$\bar{\gamma}_{i,U,t}^{R} = (1/2)(\bar{R}_t^{[2]} + \bar{R}_t^{[6]})\bar{\gamma}_{i,U,t}. \tag{27}$$

Values for the rotation-bias-adjusted continuation probabilities $\bar{\gamma}^R_{1,UU,t}$ and $\bar{\gamma}^R_{2,UU,t}$ are plotted in the last two panels of Figure 13. The objective unemployment-continuation probabilities show similar patterns to the probabilities implied by self reports, with the exception that the trend up in $\bar{p}_{2t}$, the perceived continuation-probability of the long-term unemployed, is not matched by a corresponding trend in $\bar{\gamma}^R_{2,UU,t}$, the objective probability.

6.5 Adjustments to NU transitions.[34]

We concluded in Section 6.3 that the discrepancy between reported unemployment durations and objective unemployment-continuation probabilities comes entirely from the difference between $\gamma_{2,UU}$, the objective unemployment-continuation probability for type 2 individuals, and p_2, which characterizes their perceived durations. If we are correct in our conclusion that some of the people designated by BLS as N are better regarded as U, it also means that some UN transitions are really UU continuations. What fraction ξ_{UN} of the $\gamma_{2,UN}$ transitions should we interpret to be UU continuations?

Since type 2 individuals account for 95% of those unemployed for 15 weeks and over (hereafter, U^{15+}), we look for evidence in the observed outcomes in month t of individuals who were U^{15+} in $t-2$ and N in $t-1$. Someone with a history $U^{15+}_{t-2}N_{t-1}$ has a 21.4% probability of being U^{15+} in t. We argued in Section 5 that such an individual, having been observed to be $N_{t-1}U^{15+}_t$, should be classified as U at $t-1$. This means that any $U^{15+}_{t-2}N_{t-1}U^{15+}_t$ sequence is really UUU. Thus at a minimum an average fraction $\xi_{UN} > 0.214$ of $U^{15+}_{t-2}N_{t-1}$ should be treated as UU continuations.

But $U^{15+}_{t-2}N_{t-1}$ individuals are special not just in their objective probability of returning to unemployment but also in their objective probability of successfully landing a job. Someone with a $U^{15+}_{t-2}N_{t-1}$ history has a 7.69% probability of being employed at t, far higher than usually observed for individuals classified as N_{t-1} ($P(E_t|N_{t-1}) = 4.10\%$). Suppose we view $U^{15+}_{t-2}N_{t-1}$ individuals as a mixture of two populations, with a fraction ξ_{UN} having the same employment probability in month t as someone who is reported to be $U^{15+}_{t-2}U^{15+}_{t-1}$, and the remainder with the same employment probability as someone who is truly out of the labor force in $t-1$ as represented by a history of

[34] The numbers in this section are based on an analysis of 2001:7 to 2020:2 data.

$N_{t-2}N_{t-1}$:

$$P(E_t|U_{t-2}^{15+}, N_{t-1}) = \xi_{UN} P(E_t|U_{t-2}^{15+}, U_{t-1}^{15+}) + (1 - \xi_{UN}) P(E_t|N_{t-2}, N_{t-1})$$

$$0.0769 = 0.1094\xi_{UN} + 0.0222(1 - \xi_{UN}).$$

This equation gives an estimate of $\xi_{UN} = 0.627$, which would imply an objective unemployment-continuation probability for type 2 individuals of $\gamma_{2,UU} + \xi_{UN}\gamma_{2,UN}$. We will use this estimate and look for further corroboration of what it implies for other variables in the analysis below.

6.6 Implications for labor-force participation and unemployment rates.

We concluded in Section 5 that $m_{N,t-1}^{\sharp}$ should be added to the fraction of the sample deemed to be unemployed on the basis of observed $N_{t-1}U_t^{5+}$ transitions. The recommendation in Section 6.5 was that we should further add $m_{N,t-1}^{\flat}$ to correct for the fraction of $U_{t-2}^{15+}N_{t-1}$ transitions that likely represent continuations of long-term unemployment, where $m_{N,t-1}^{\flat}$ is calculated from

$$m_{N,t-1}^{\flat} = \pi_{U,t-1}^{*} \bar{w}_{2,t-1} \bar{\gamma}_{2,UN,t-1}^{R} \xi_{UN}.$$

Here $\pi_{U,t-1}^{*}$ is the fourth element of π_{t-1}^{*}, $\bar{w}_{2,t-1}$ and $\bar{\gamma}_{2,UN,t-1}^{R}$ are the parameter estimates described in Section 6.4, and we fix $\xi_{UN} = 0.627$ at the full-sample average[35].

The adjustments $m_{N,t-1}^{\sharp}$ and $m_{N,t-1}^{\flat}$ entail some double-counting because individuals who are $U_{t-2}^{15+}N_{t-1}U_t^{5+}$ would be included in both $m_{N,t-1}^{\sharp}$ and $m_{N,t-1}^{\flat}$. We correct for this by calculating $k^{\natural}$, the fraction of $m_{Nt}^{\sharp} + m_{Nt}^{\flat}$ that comes from double-counting the same individuals. For the full sample that fraction is

$$k^{\natural} = \frac{m_N^{\natural}}{m_N^{\sharp} + m_N^{\flat}} = \frac{0.0006}{0.0037 + 0.0026} = 0.095.$$

This results in the monthly estimate $m_{Nt}^{\natural} = k^{\natural}(m_{Nt}^{\sharp} + m_{Nt}^{\flat})$. Our final estimates that correct for

[35] We obtained similar results allowing $\xi_{UN,t}$ to change over time.

rotation bias, non-randomly missing observations, and misclassified N are then

$$\begin{bmatrix} \pi^{RMN}_{E,t-1} \\ \pi^{RMN}_{N,t-1} \\ \pi^{RMN}_{M,t-1} \\ \pi^{RMN}_{U,t-1} \end{bmatrix} = \begin{bmatrix} \pi^{*}_{E,t-1} + \pi^{*}_{M,t-1}\overline{m}_{E,t-1} \\ \pi^{*}_{N,t-1} + \pi^{*}_{M,t-1}\overline{m}_{N,t-1} - m^{\sharp}_{N,t-1} - m^{\flat}_{N,t-1} + m^{\natural}_{N,t-1} \\ \pi^{*}_{M,t-1}(1 - \overline{m}_{E,t-1} - \overline{m}_{N,t-1} - \overline{m}_{U,t-1}) \\ \pi^{*}_{U,t-1} + \pi^{*}_{M,t-1}\overline{m}_{U,t-1} + m^{\sharp}_{N,t-1} + m^{\flat}_{N,t-1} - m^{\natural}_{N,t-1} \end{bmatrix}. \tag{28}$$

Our combined adjustments to the unemployment rate and labor-force participation rate are then

$$\tilde{u}^{RMN}_t = \pi^{RMN}_{U,t}/(\pi^{RMN}_{E,t} + \pi^{RMN}_{U,t}) - \pi_{U,t}/(\pi_{E,t} + \pi_{U,t})$$

$$\tilde{\ell}^{RMN}_t = \left(\pi^{RMN}_{E,t} + \pi^{RMN}_{U,t}\right) / \left(\pi^{RMN}_{E,t} + \pi^{RMN}_{N,t} + \pi^{RMN}_{U,t}\right) - (\pi_{E,t} + \pi_{U,t})/(\pi_{E,t} + \pi_{N,t} + \pi_{U,t}).$$

The average values of these adjustments over our sample are summarized in Table 6. Altogether the adjustments add 2% on average to the unemployment rate and 2.1% to the labor-force participation rate. The biggest single factor in the adjustments comes from $m^{\sharp}_{N,t-1}$ as explained in Section 5.

Seasonally adjusted estimates for each month are plotted in Figure 7. As documented in Sections 3-5, each of the adjustments is countercyclical – rotation bias is highest, MX observations are more likely to be MU, and $N_{t-1}U^{5+}_t$ are all biggest when the BLS-reported unemployment rate is higher. The combined effect of all three adjustments is that BLS may have underestimated the unemployment rate by 2.6% and the labor-force participation rate by 3.6% in May 2020. Our adjustment to the labor-force participation rate also shows a modest increasing trend over time, rising from 1.5% in 2001:7 to 2.6% in 2020:11. In other words, the usual statistics may have overstated the magnitude of the decline in labor-force participation over this period by 1.1%.

Note that our method does not allow for the possibility that some reported E are truly U. In normal times we feel this is the correct conservative approach. However, the unusual challenges to collecting data during the pandemic led the BLS to conclude that as much as 5% of the labor-force in April 2020 were incorrectly counted as E (missing work for "other reasons") whereas according to BLS guidelines those individuals should have been counted as U on temporary layoff. Thus as large as our corrections to the unemployment rate are, the true rate in the spring of 2020 was likely much higher.

The last column of Table 6 shows that while rotation bias matters for the employment-population ratio, the ratio is unchanged after further correcting for missing observations or misclassified N. Thus apart from the special circumstances in the spring of 2020, the employment-population ratio could be a more robust measure of labor-market slack in the presence of increasing nonresponses and errors in responses in the CPS.

6.7 Implications for estimated transition probabilities.

Our concept for calculating transition probabilities is that used by Fujita and Ramey (2009) and Elsby, Hobijn and Şahin (2010) – we base our estimates on the observed month-to-month transitions of individuals. However, our estimates differ from theirs in that we correct for rotation bias, nonrandom missing observations, and misclassified N.

Our approach interprets fractions $\bar{m}_{E,t}, \bar{m}_{N,t}$ and $\bar{m}_{U,t}$ of the $U_{t-1}M_t$ transitions as UE, UN, and UU transitions, respectively, and interprets a fraction ξ_{UN} of the type 2 UN transitions as UU continuations. Transition probabilities that correct for rotation bias, nonrandom missing observations, and misclassified N are thus given by

$$\bar{\gamma}^{RMN}_{1,U,t} = \begin{bmatrix} \bar{\gamma}^R_{1,UE,t} + \bar{\gamma}^R_{1,UM,t}\bar{m}_{E,t} \\ \bar{\gamma}^R_{1,UN,t} + \bar{\gamma}^R_{1,UM,t}\bar{m}_{N,t} \\ \bar{\gamma}^R_{1,UM,t}(1 - \bar{m}_{E,t} - \bar{m}_{N,t} - \bar{m}_{U,t}) \\ \bar{\gamma}^R_{1,UU,t} + \bar{\gamma}^R_{1,UM,t}\bar{m}_{U,t} \end{bmatrix}$$

$$\bar{\gamma}^{RMN}_{2,U,t} = \begin{bmatrix} \bar{\gamma}^R_{2,UE,t} + \bar{\gamma}^R_{2,UM,t}\bar{m}_{E,t} \\ (1-\xi_{UN})\bar{\gamma}^R_{2,UN,t} + \bar{\gamma}^R_{2,UM,t}\bar{m}_{N,t} \\ \bar{\gamma}^R_{2,UM,t}(1 - \bar{m}_{E,t} - \bar{m}_{N,t} - \bar{m}_{U,t}) \\ \bar{\gamma}^R_{2,UU,t} + \bar{\gamma}^R_{2,UM,t}\bar{m}_{U,t} + \xi_{UN}\bar{\gamma}^R_{2,UN,t} \end{bmatrix}.$$

To calculate the true unemployment-continuation probability for type i workers between t and $t+1$, we assume that whether an individual becomes an unallocated missing observation is independent of status:

$$\tilde{\gamma}^*_{i,UU,t} = \bar{\gamma}^{RMN}_{i,UU,t}/(1 - \bar{\gamma}^{RMN}_{i,UM,t}).$$

The estimate $\tilde{\gamma}^*_{2,UU}$ averages 0.76, well below $p_2^{4.33} = 0.89$, the value we would have expected based on reported unemployment durations. Nevertheless, the adjustment goes a fair way toward reconciling perceived durations with objective continuation probabilities. One source of the remaining discrepancy between our estimate of the objective continuation probability $\tilde{\gamma}^*_{2,UU}$ and the perceived duration of job search p_2 is on-the-job search. Recall from Section 6.2 that EU^{5+} transitions account for 26% of EU observations, with many EU individuals reporting duration longer than 6 months. As noted by Kudlyak and Lange (2018), we could interpret these individuals as correctly reporting how long they have been looking for a job or looking for a better job, while still defending the estimate $\tilde{\gamma}^*_{2,UU}$ as a correct summary of the true probability of remaining unemployed without an intervening spell of employment. A second possible source of discrepancy between $\tilde{\gamma}^*_{2,UU}$ and p_2 is that individuals are reporting not the length of a continuous spell of unemployment but instead how long it has been since their last good job (Elsby et al. (2011); Farber and Valletta (2015)). We conclude that our procedure of adjusting unemployment-continuation probabilities up, but not all the way to those implied by reported job-search durations, is the correct way to reconcile the data.

6.8 Average unemployment-continuation probabilities.

We can also calculate the fraction $\tilde{w}_{i,t-1}$ of total unemployed individuals that are of type i. Consider the last row of equation (28). For the first term in that equation ($\pi^*_{U,t-1}$), we know the fraction of type i from the estimate of $\bar{w}_{i,t-1}$. We assume the same fraction $\bar{w}_{i,t-1}$ could be used to impute types for the second term ($\pi^*_{M,t-1}\bar{m}_{U,t-1}$). The third term ($m^\sharp_{N,t-1}$) is derived from observed NU^{5+} transitions, for which we have estimated the fraction of type 1 to be $\bar{q}_{5,NU,t-1}/(\bar{q}_{5,NU,t-1} + \bar{q}_{6,NU,t-1})$. The last two terms by construction come solely from type 2 individuals. We thus estimate

$$\tilde{w}_{1,t} = \frac{\bar{w}_{1,t}(\pi^*_{U,t} + \pi^*_{M,t}\bar{m}_{U,t}) + m^\sharp_{N,t}\bar{q}_{5,NU,t}/(\bar{q}^R_{5,NU,t} + \bar{q}^R_{6,NU,t})}{\pi^{RMN}_{U,t}}$$

and $\tilde{w}_{2,t} = 1 - \tilde{w}_{1,t}$. Our estimate of the true monthly continuation probability averaged across all individuals who are truly unemployed is then

$$\tilde{w}_{1,t}\tilde{\gamma}^*_{1,UU,t} + \tilde{w}_{2,t}\tilde{\gamma}^*_{2,UU,t}.$$

This is the series plotted as the blue dotted line in Panel A of Figure 1.

6.9 Implications for estimates of new flows into unemployment and unemployment duration.

The corrections above imply that a fraction $\tilde{w}_{i,t}\pi^{RMN}_{U,t}/(1-\pi^{RMN}_{M,t})$ of individuals in the sample are truly unemployed of type $i \in \{1,2\}$ in month t. Of these, a fraction $\tilde{\gamma}^*_{i,UU,t+1}$ are still unemployed the next month, giving rise to

$$\tilde{V}_{i,t+1} = \frac{\tilde{w}_{i,t+1}\pi^{RMN}_{U,t+1}}{(1-\pi^{RMN}_{M,t+1})} - \frac{\tilde{\gamma}^*_{i,UU,t+1}\tilde{w}_{i,t}\pi^{RMN}_{U,t}}{(1-\pi^{RMN}_{M,t})} \tag{29}$$

as an estimate of the number of individuals of type i who are newly unemployed in month $t+1$ and $\tilde{V}_{t+1} = \tilde{V}_{1,t+1} + \tilde{V}_{2,t+1}$ as the total number of newly unemployed. This is the series that was plotted as the dotted blue line in Panel B of Figure 1.

Let $\tilde{V}_{i,t-d+1}$ denote the number of newly unemployed of type i at $t-d+1$ as calculated in (29). A fraction $\tilde{\gamma}^*_{i,UU,t-d+2}$ will still be unemployed at $t-d+2$. Thus the number unemployed for exactly d months as of month t would be given by

$$\tilde{U}^d_{i,t} = \tilde{V}_{i,t-d+1}\tilde{\gamma}^*_{i,UU,t-d+2}\cdots\tilde{\gamma}^*_{i,UU,t-2}\tilde{\gamma}^*_{i,UU,t-1}\tilde{\gamma}^*_{i,UU,t}. \tag{30}$$

This implies an average unemployment duration of those who are unemployed in month t of

$$\tilde{d}_t = \frac{\sum_{d=1}^{48} d(\tilde{U}^d_{1,t} + \tilde{U}^d_{2,t})}{\sum_{d=1}^{48}(\tilde{U}^d_{1,t} + \tilde{U}^d_{2,t})}. \tag{31}$$

Dividing the average monthly duration in (31) by 4.33 gives the unemployment duration in weeks plotted as the blue line in Panel E of Figure 1. Our series is much lower on average and less cyclically variable than the BLS measure in black. Nevertheless, our conclusion that the surge of newly unemployed in April 2020 brought the average duration down to about 6 weeks is the same as implied by the BLS series. We calculated the average values of (30) over all months t in our sample and report in Table 11 the average percentage of the truly unemployed for whom the true duration is less than 5 weeks (1 month), 5-14 weeks (2-3 months), 15-26 weeks (4-6 months) and longer than 26 months (7 months and over), along with the average duration. Our estimate of the average duration of unemployment is only 15 weeks, about 11 weeks lower than the BLS reports. Kudlyak and Lange (2018) constructed estimates of the number of newly unemployed as a fraction of total unemployed by (1) counting all $E_{t-1}U_t$ as newly unemployed despite the duration of search reported at t, and (2) also counting all $N_{t-1}U_t$ as newly unemployed. Our estimate of the fraction of individuals unemployed for less than 5 weeks, 36.7%, is in between their two estimates (29.1% and 46.1%, respectively) because we designate some, but not all, of the $N_{t-1}U_t$ as unemployed at $t-1$. Their two methods produced estimates of 37.5% and 24.1%, respectively, for the fraction of unemployed with duration greater than 14 weeks, with our estimate of 32% again in between those two. Although their approach did not allow them to uncover the average duration of unemployment, their calculations support our conclusion that the BLS estimates substantially overstate the number of long-term unemployed.

7 Comparison with other estimates.

A number of other researchers have made efforts to address some of the problems addressed by our paper. In this section we compare our estimates with other approaches.

Unemployment rate. One method that some researchers have relied on to adjust the data is based on follow-up interviews conducted by the BLS a half-century ago. Abowd and Zellner (1986, Table 6) found that 9.5% of individuals who were determined on the basis of a second interview to have been truly unemployed were counted as not in the labor force based on answers given in the first interview. Poterba and Summers (1986 , Table II) put the number at 13.6%. Researchers like Elsby, Hobijn, and Şahin (2015) have assumed that the Abowd-Zellner proportions have continued to hold for all subsequent observations. We have extended their estimates through 2020, shown as the dotted green line in the first panel of Figure 14. These adjusted estimates are approximately equal to 1.13 times the measured unemployment rate minus a few tenths of a percent.[36] They thus share the countercyclical property of our adjustments – when the measured unemployment rate is higher, the number of percentage points by which the BLS measure understates the true unemployment rate is bigger. However, when the unemployment rate dropped to extremely low levels in 2019, the Abowd-Zellner adjustment becomes negligible. Note that although Abowd and Zellner also studied the bias arising from nonrandom missing observations, those corrections are not included in the extension of Abowd-Zellner as commonly applied, and there is no correction for rotation bias. Feng and Hu (2013, page 1055) raise a number of other concerns about the reliability of second interviews and the invariance of these tendencies over time.

Other researchers like Biemer and Bushery (2000), Feng and Hu (2013), and Shibata (2019) assumed that the reported data differ from latent true values, with identification coming from assumptions about the joint dynamics of the true values and measurement error. We updated the estimates using Feng and Hu's algorithm in the solid red line in Figure 14A. Although derived under very different assumptions from ours, the two series are quite similar up to 2008. The Feng-Hu approach leads to an even higher estimate of the unemployment rate at the peak of the two recessions. This is due to the fact that their algorithm reclassifies some E as U whereas our approach does not make any subtractions from measured E. Shibata (2019) raised concerns about implausible transition probabilities implied by the Feng-Hu algorithm, whereas our approach is based on an integrated reconciliation of stocks and flows. Our approach also explains well the non-Markov predictability of labor-force status documented by Kudlyak and Lange (2018).

Recently the Federal Reserve Bank of New York has added detailed questions to their Survey of Consumer Expectations about an individual's search effort, search methods and outcomes, and the incidence of informal recruiting methods. Faberman et al. (2019) found that if one defines unemployment to mean someone who actively searched and is available for work , the unemployment

[36] Let $\hat{\pi}_{E,t}, \hat{\pi}_{U,t}, \hat{\pi}_{N,t}$ denote the initial reported values as fractions of the (E_t, U_t, N_t) sample and $\tilde{\pi}^{AZ}_{E,t}, \tilde{\pi}^{AZ}_{U,t}, \tilde{\pi}^{AZ}_{N,t}$ the corrected values. Expression (1) in Elsby, Hobijn, and Şahin (2015) states

$$\begin{bmatrix} \hat{\pi}_{E,t} \\ \hat{\pi}_{U,t} \\ \hat{\pi}_{N,t} \end{bmatrix} = \begin{bmatrix} 0.9878 & 0.0191 & 0.0050 \\ 0.0018 & 0.8857 & 0.0029 \\ 0.0103 & 0.0952 & 0.9921 \end{bmatrix} \begin{bmatrix} \tilde{\pi}^{AZ}_{E,t} \\ \tilde{\pi}^{AZ}_{U,t} \\ \tilde{\pi}^{AZ}_{N,t} \end{bmatrix}$$

which implies

$$\begin{bmatrix} \tilde{\pi}^{AZ}_{E,t} \\ \tilde{\pi}^{AZ}_{U,t} \\ \tilde{\pi}^{AZ}_{N,t} \end{bmatrix} = \begin{bmatrix} 1.0124 & -0.0213 & -0.0050 \\ -0.0020 & 1.1294 & -0.0033 \\ -0.0103 & -0.1082 & 1.0083 \end{bmatrix} \begin{bmatrix} \hat{\pi}_{E,t} \\ \hat{\pi}_{U,t} \\ \hat{\pi}_{N,t} \end{bmatrix}.$$

rate in the U.S. over October 2013 to December 2017 would have been 1.7% higher on average than the figures reported by the BLS. This is close to the figure implied by our final adjustment, which is 2.1% higher than the BLS figure over this period. Faberman et al.'s measure does not account for nonrandom missing observations, which could explain the 0.4% difference between their estimate and ours.

Figure 14B compares our adjusted estimate $\tilde{u}_t$ (in dotted blue) with three different unemployment rates reported by the BLS– the usual U3 unemployment rate (black) along with U5 unemployment (dotted green), which includes discouraged workers and all other marginally attached workers, and U6 unemployment (green) which adds people who are employed part-time for economic reasons. Our adjustment includes more individuals than U5, but far less than U6.

New flows into unemployment. Figure 14C compares our estimate of new flows into unemployment (shown in dotted blue) with several others. Researchers like Fujita and Ramey (2012) based their calculation on the number of EU and NU transitions among those with two consecutive months of nonmissing observations,

$$\hat{V}_t = \frac{\sum_{j\in J}\left(y^{[j]}_{E,U,t} + y^{[j]}_{N,U,t}\right)}{\sum_{j\in J}\left(y^{[j]}_{E,E,t} + y^{[j]}_{E,N,t} + y^{[j]}_{E,U,t} + y^{[j]}_{N,E,t} + y^{[j]}_{N,N,t} + y^{[j]}_{N,U,t} + y^{[j]}_{U,E,t} + y^{[j]}_{U,N,t} + y^{[j]}_{U,U,t}\right)}, \tag{32}$$

shown as the solid black line. Our estimate is always above $\hat{V}_t$. One important reason is rotation bias, which causes flows into unemployment as calculated from the numerator of (32) to be smaller than flows out of unemployment even in months when the measured unemployment rate is constant or even rising. One can see the effect of rotation bias by replacing $\sum_{j\in J} y^{[j]}_{X_1,X_2,t}$ in (32) by the estimate $\pi^*_{X_1,t-1}\pi^*_{X_1,X_2,t}$. This corrects the calculation for rotation bias but makes no other adjustments. The resulting series $\hat{V}^*_t$ is shown as the turquoise line in Figure 14C. This turns out to be quite similar to the BLS adjusted flow series (dashed green). These flows are inconsistent with the magnitude of the increase in unemployment that we believe characterized both recessions, and the BLS flows continued through the end of 2020 to underestimate the number of people newly unemployed each month. The Abowd-Zellner adjusted flows into unemployment in the Great Recession and years after are considerably below even the direct Fujita estimates.

Unemployment-continuation probability. Figure 14D plots alternative estimates of the probability that someone unemployed this month will still be unemployed the following month. The BLS flows series are modestly above the direct measure based on a calculation like (32) but well below the adjustments that we believe are warranted. They make no adjustment for misclassified UN transitions and do not fully account for either rotation bias or nonrandom missing observations. The Abowd-Zellner continuation probabilities turn out to be similar to ours, though AZ attribute more of the rise in unemployment during recessions to continuing unemployment and less to new inflows than we do.

Summary. Other researchers have approached the issues in our paper from a variety of perspectives. Although the methods differ widely, two conclusions are universal: the BLS measures underestimate the number of people who are unemployed, and the magnitude of the bias is larger when the true unemployment rate is higher.

8 Conclusion.

The data underlying the CPS contain multiple internal inconsistencies. These include the facts that people's answers change the more times they are asked the same question, stock estimates are inconsistent with flow estimates, missing observations are not random, reported unemployment durations are inconsistent with reported labor-force histories, and people prefer to report some numbers over others. Ours is the first paper to attempt a unified reconciliation of these issues. We conclude that the U.S. unemployment rate and labor-force continuation rates are higher than conventionally reported while the average duration of unemployment is considerably lower.

References

Abowd, John M., and Arnold Zellner (1985). "Estimating Gross Labor-Force Flows." *Journal of Business and Economic Statistics* 3, no. 3: 254-283.

Abraham, Katharine G., and Shimer Robert (2001). "Changes in Unemployment Duration and Labor Force Attachment." NBER Working Paper 8513.

Ahn, Hie Joo, and James D. Hamilton (2020). "Heterogeneity and Unemployment Dynamics." *Journal of Business and Economic Statistics,* forthcoming.

Ahn, Hie Joo, and Ling Shao (2021). "The Cyclicality of On-the-Job Search Effort." *The B.E. Journal of Macroeconomics,* 21(1): 185-220.

Bailar, Barbara A (1975). "The Effects of Rotation Group Bias on Estimates from Panel Surveys." *Journal of the American Statistical Association* 70: 23-30.

Baker, Michael (1992). "Digit Preference in CPS Unemployment Data." *Economics Letters*, 39(1):117-121.

Bureau of Labor Statistics (2020). "Employment Situation Summary," May 8 (https://www.bls.gov/news.release/archives/empsit_05082020.htm).

Biemer, Paul P., and John M. Bushery (2000). "On the Validity of Markov Latent Class Analysis for Estimating Classification Error in Labor Force Data." *Survey Methodology* 26: 139-152.

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 (2015). "On the Importance of the Participation Margin for Labor Market Fluctuations." *Journal of Monetary Economics* 72: 64-82.

Elsby, Michael WL, 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.

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, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa (2019). "Job Search Behavior among the Employed and Non-Employed." working paper, University of Texas at Austin.

Farber, Henry S., and Robert G. Valletta (2015). "Do Extended Unemployment Benefits

Lengthen Unemployment Spells? Evidence from Recent Cycles in the US Labor Market." *Journal of Human Resources* 50: 873-909.

Feng, Shuaizhang (2001). "The Longitudinal Matching of Current Population Surveys: A Proposed Algorithm." *Journal of Economic and Social Measurement* 27: 71-91.

Feng, Shuaizhang, and Yingyao Hu (2013). "Misclassification Errors and the Underestimation of the US Unemployment rate." *American Economic Review* 103: 1054-70.

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

Hall, Robert E., and Marianna Kudlyak (2019). "Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics." NBER Working Paper 25625.

Halpern-Manners, Andrew, and John Robert Warren (2012). "Panel Conditioning in Longitudinal Studies: Evidence from Labor Force Items in the Current Population Survey." *Demography* 49(4): 1499-1519.

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

Hansen Morris H., William N. Hurwitz, Harold Nisselson, and Joseph Steinberg (1955). "The Redesign of the Census Current Population Survey." *Journal of the American Statistical Association* 50: 701-719.

Hirsch, Barry T., and John V. Winters (2016). "Rotation Group Bias in Measures of Multiple Job Holding." *Economics Letters* 147: 160–163.

Ilg, Randy (2011). "How Long before the Unemployed Find Jobs or Quit Looking?" *BLS Issues in Labor Statistics* May: 1-6.

Krueger, Alan B., Alexandre Mas, and Xiaotong Niu (2017). "The Evolution of Rotation Group Bias: Will the Real Unemployment Rate Please Stand Up?" *Review of Economics and Statistics* 99: 258-264.

Kudlyak, Marianna, and Fabian Lange (2018). "Measuring Heterogeneity in Job Finding Rates Among the Nonemployed Using Labor Force Status Histories." Working paper, Federal Reserve Bank of San Francisco.

Madrian, Brigitte C., and Lars John Lefgren (2000). "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents." *Journal of Economic and Social Measurement* 26(1): 31-62.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan (2015). "Household Surveys in Crisis." *Journal of Economic Perspectives*, 29(4): 199-226.

Nekarda, Christopher J. (2009). "A Longitudinal Analysis of the Current Population Survey: Assessing the Cyclical Bias of Geographic Mobility." Federal Reserve Board of Governors.

Poterba, James M. and Lawrence H. Summers (1986). "Reporting Errors and Labor Market Dynamics." *Econometrica* 54: 1319-1338.

Rothstein, Jesse (2011). "Unemployment Insurance and Job Search in the Great Recession." *Brookings Papers on Economic Activity*, Fall 2011: 143-196.

Ryu, Hang K., and Daniel J. Slottje (2000). "Estimating the Density of Unemployment Duration Based on Contaminated Samples or Small Samples." *Journal of Econometrics*, 95(1): 131-156.

Shibata, Ippei (2019). "Reassessing Classification Errors in the Analysis of Labor Market Dynamics." IMF Working Paper.

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

Silverstone, Brian, and Will Bel (2010). "Labour Market Flows in New Zealand: Some Questions and Some Answers." 51st Conference of the New Zealand Association of Economists, Auckland, vol. 30.

Solon, Gary (1986). "Effects of Rotation Group Bias on Estimation of Unemployment." *Journal of Business and Economic Statistics* 4: 105-109.

Torelli, Nicola, and Ugo Trivellato (1993). "Modelling Inaccuracies in Job-search Duration Data." *Journal of Econometrics* 59(1-2): 187-211.

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 Brakel, Jan A., and Sabine Krieg (2015). "Dealing with Small Sample Sizes, Rotation Group Bias and Discontinuities in a Rotating Panel Design." *Survey Methodology* 41: 267-296.

Table 1. Average numbers of individuals with indicated status across different rotation groups, July 2001 to December 2020.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
rotation	E	N	M	U	total	U/(U+E)	(U+E)/(U+E+N)
1	7,727	4,320	5,678	552	18,277	6.7	65.7
2	7,887	4,545	5,317	541	18,289	6.4	65.0
3	7,906	4,599	5,272	524	18,301	6.2	64.7
4	7,897	4,622	5,279	513	18,312	6.1	64.5
5	7,693	4,571	5,662	499	18,426	6.1	64.2
6	7,810	4,662	5,471	492	18,436	5.9	64.0
7	7,844	4,679	5,435	485	18,443	5.8	64.0
8	7,898	4,710	5,356	487	18,451	5.8	64.0

Table 2. Values of rotation-group bias parameters for full sample, July 2001 to December 2020.

j	1	2	3	4	5	6	7	8
$\theta_{EM}^{[j]}$	0	0.0196	0.0213	0.0196	0	0.0021	0.0059	0.0121
$\theta_{NU}^{[j]}$	0	0.0025	0.0063	0.0086	0.0125	0.0138	0.0155	0.0148
$\theta_{NM}^{[j]}$	0	0.0463	0.0532	0.0550	0.0346	0.0515	0.0529	0.0590

Notes to Table 2. Calculated from the expressions in footnote 7 for $\pi_X^{[j]} = \frac{\sum_{t=1}^T y_{X,t}^{[j]}}{\sum_{t=1}^T (y_{E,t}^{[j]}+y_{N,t}^{[j]}+y_{M,t}^{[j]}+y_{U,t}^{[j]})}$.

Table 3. Estimated average fractions of individuals π_X^* who would have reported labor status *E, N, M*, *U* and transition probabilities π_{X_1,X_2}^* if all individuals were being interviewed for the first time, August 2001 to December 2020.

$$\begin{bmatrix} \pi_E^* \\ \pi_N^* \\ \pi_M^* \\ \pi_U^* \end{bmatrix} = \begin{bmatrix} 0.4213 \\ 0.2365 \\ 0.3121 \\ 0.0301 \end{bmatrix} \qquad \begin{bmatrix} \pi_{EE}^* & \pi_{NE}^* & \pi_{ME}^* & \pi_{UE}^* \\ \pi_{EN}^* & \pi_{NN}^* & \pi_{MN}^* & \pi_{UN}^* \\ \pi_{EM}^* & \pi_{NM}^* & \pi_{MM}^* & \pi_{UM}^* \\ \pi_{EU}^* & \pi_{NU}^* & \pi_{MU}^* & \pi_{UU}^* \end{bmatrix} = \begin{bmatrix} 0.8958 & 0.0361 & 0.0935 & 0.2051 \\ 0.0257 & 0.8664 & 0.0474 & 0.1983 \\ 0.0658 & 0.0689 & 0.8502 & 0.0905 \\ 0.0127 & 0.0286 & 0.0089 & 0.5060 \end{bmatrix}$$

Table 4. Characteristics of *U* and *N* as a function of rotation.

	1	2	3	4	5	6	7	8	avg(2-8)
(1) *N*/(*E*+*N*+*U*)	34.2	34.9	35.2	35.4	35.7	35.8	35.9	35.8	35.5
(2) retired/(*E*+*N*+*U*)	15.5	16.2	16.4	16.6	16.2	16.6	16.8	16.8	16.5
(3) disabled/((*E*+*N*+*U*)	4.6	5.0	5.2	5.3	4.9	5.2	5.3	5.4	5.2
(4) Probability *E* or *U* in *j* +1 given retired or disabled in *j*	1.75	1.90	1.87		1.91	1.82	1.81		1.86
(5) Standard error	(0.017)	(0.017)	(0.017)		(0.018)	(0.017)	(0.017)		(0.006)
(6) *U* (self-report)/(*E*+*U*)	3.3	3.1	3.0	2.9	3.0	2.9	2.8	2.8	2.9
(7) *U* (proxy)/(*E*+*U*)	3.2	3.1	3.0	3.0	3.0	2.9	2.9	2.9	3.0
(8) *E*/(*E*+*U*+*N*)	61.5	60.9	60.8	60.8	60.4	60.3	60.4	60.4	60.6
(9) $\dot{E}$/(*E*+*U*+*N*)	16.6	16.1	15.9	15.8	15.6	15.5	15.5	15.5	15.7
(10) *U*/(*E*+*U*)	6.6	6.4	6.2	6.0	6.0	5.9	5.8	5.7	
(11) *U*/(*E*+*U*) given M1		7.4	6.8	6.3	5.6	5.4	5.2	5.1	
(12) *U*/(*E*+*U*) given M1 and M2			8.8	8.1	6.5	6.3	5.9	5.9	
(13) *U*/(*E*+*U*) given M1-M3				9.9	6.9	6.7	6.3	6.5	
(14) *U*/(*E*+*U*) given M1-M4					9.3	8.6	7.9	7.6	
(15) *U/(U+E)* given M1-M5						9.2	8.5	7.8	
(16) *U/(U+E)* given M1-M6							10.2	9.0	
(17) Total *N* (in thousands)	976	1024	1035	1038	1028	1047	1051	1057	

Notes to Table 4. Row (1): *N* as a percent of *E+N+U*. (2): retired individuals as a percent of *E+N+U*. (3): disabled individuals as a percent of *E+N+U*. (4): probability that an individual who is retired or disabled in rotation *j* will be *E* or *U* in rotation *j* + 1. (5): standard error of row (4). (6): individuals who report their own status to be *U* as a percent of the labor force. (7): individuals whose status is reported by another member of the household to be *U* as a percent of the labor force. (8): *E* as a percent of *E+N+U*. (9): Part-time unemployed plus those usually employed but not employed this week as a percent of *E+N+U*. (10): unemployment rate as a function of rotation among individuals who are not missing in rotation 1. (11): unemployment rate as a function of rotation among individuals who are missing in rotation 1 but not missing in rotation 2. (12): unemployment rate among individuals who are missing in rotations 1 and 2 but not missing in 3. Rows (13)-(16): unemployment rate among individuals who are missing in rotations 1 through *j* – 1 but not missing in *j.* (17): Total number of individuals counted as not in the labor force from each rotation. All numbers are reported as percent except for last row which is in thousands of individuals. Rows (1)-(9) and (17) refer to average over Jul 2001 to Feb 2020 while rows (8)-(14) are over Sep 2002 to Feb 2020.

Table 5. Unemployment rates in rotation *j* and *j* + 1 among individuals who are not missing in either *j* or $j+1$, July 2001 to February 2020.

Rotation	u_j	u_{j+1}	difference
$j=1$	6.50	6.23	0.27
$j=2$	6.22	5.99	0.24
$j=3$	6.00	5.85	0.15
$j=5$	5.91	5.71	0.19
$j=6$	5.73	5.58	0.15
$j=7$	5.62	5.59	0.04

Table 6. Effects of adjustments on average reported unemployment and labor-force participation rates, July 2001 to December 2020.

	Unemployment rate	Labor-force participation rate	Employment-population ratio
Unadjusted BLS	6.1%	64.4%	60.5%
Corrected for rotation-group bias only	6.7%	65.5%	61.1%
Corrected for rotation-group bias and missing observations	7.0%	65.8%	61.1%
Corrected for rotation-group bias, missing observations, and long-term unemployed	8.1%	66.5%	61.1%

Table 7. Month *t* + 1 employment probabilities for *UUU* and *UNU* histories, July 2001 to February 2020.

UUU	Probability	*UNU*	Probability
$U_{t-2}^{1.4}, U_{t-1}^{5.14}, U_t^{5.14}$	0.19	$U_{t-2}^{1.4}, N_{t-1}, U_t^{5.14}$	0.15
$U_{t-2}^{5.14}, U_{t-1}^{5.14}, U_t^{15.26}$	0.16	$U_{t-2}^{5.14}, N_{t-1}, U_t^{15.26}$	0.15
$U_{t-2}^{15.26}, U_{t-1}^{15.26}, U_t^{15.26}$	0.14	$U_{t-2}^{15.26}, N_{t-1}, U_t^{15.26}$	0.14
$U_{t-2}^{15.26}, U_{t-1}^{27+}, U_t^{27+}$	0.12	$U_{t-2}^{15.26}, N_{t-1}, U_t^{27+}$	0.10
$U_{t-2}^{27+}, U_{t-1}^{27+}, U_t^{27+}$	0.08	$U_{t-2}^{27+}, N_{t-1}, U_t^{27+}$	0.07

Table 8. Parameters estimated separately for rotation 1, rotation 5, and NX, EX and MX transitions from rotation 1 to rotation 2, July 2001 to December 2020.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
param	rotation 1 only	std error	rotation 5 only	std error	NX only	std error	EX only	std error	MX only	std error
p_1	0.8293	0.0044	0.8314	0.0053	0.7528	0.0103	0.7238	0.0286	0.8348	0.0067
p_2	0.9732	0.0025	0.9728	0.0026	0.9740	0.0023	0.9672	0.0037	0.9728	0.0028
w_1	0.4344	0.0464	0.4139	0.0491						
π_E	0.4228		0.4175							
π_N	0.2364	0.0049	0.2481	0.0050						
π_M	0.3106	0.0053	0.3073	0.0057						
π_U	0.0302	0.0029	0.0271	0.0026						
π_{XE}					0.0383	0.0015	0.8858	0.0059	0.1332	0.0076
π_{XN}					0.8746	0.0020	0.0321	0.0006	0.0690	0.0052
π_{XM}					0.0626		0.0687		0.7869	
π_{XU}					0.0246	0.0020	0.0134	0.0006	0.0109	0.0006
q_1					0.0917		0.2023		0.0830	
q_2					0.0775	0.0054	0.1974	0.0088	0.0982	0.0069
q_3					0.0820	0.0055	0.1701	0.0068	0.0804	0.0053
q_4					0.0529	0.0029	0.1307	0.0068	0.0829	0.0048
q_5					0.1896	0.0193	0.1245	0.0070	0.2338	0.0258
q_6					0.5063	0.0393	0.1751	0.0127	0.4218	0.0430
$q_5 + q_6$					0.6959		0.2996		0.6556	
$\theta_{A,1}$	0.1240	0.0037	0.1302	0.0061	0.2099	0.0287	0.0607	0.0336	0.0466	0.0223
$\theta_{A,2}$	0.7668	0.0077	0.7277	0.0100	0.7448	0.0101	0.6893	0.0314	0.7286	0.0132
$\theta_{A,3}$	0.4783	0.0090	0.4536	0.0075	0.4953	0.0188	0.3663	0.0232	0.4886	0.0266
$\theta_{A,4}$	0.9275	0.0037	0.8746	0.0074	0.8628	0.0127	0.8376	0.0274	0.8583	0.0095
$\theta_{A,5}$	0.7028	0.0170	0.6835	0.0088	0.7010	0.0160	0.6866	0.0305	0.7432	0.0321
$\theta_{A,6}$	0.9317	0.0075	0.9038	0.0047	0.8753	0.0137	0.8250	0.0151	0.8807	0.0170
$\theta_{A,7}$	0.9366	0.0168	0.9533	0.0020	0.9511	0.0138	0.9736	0.0093	0.9366	0.0157
$\theta_{A,8}$	0.8988	0.0074	0.8438	0.0165	0.7246	0.0345	0.7454	0.0260	0.8001	0.0281
$\theta_{A,9}$	0.8930	0.0520	0.9405	0.0069	0.9514	0.0141	0.9486	0.0366	0.9364	0.0125
$\theta_{A,10}$	0.1622	0.0071	0.1467	0.0052	0.1990	0.0095	0.1310	0.0094	0.1417	0.0069
$\theta_{A,11}$	0.4883	0.0085	0.4932	0.0065	0.5801	0.0130	0.4937	0.0212	0.4770	0.0211
$\theta_{A,12}$	0.9230	0.0155	0.8882	0.0134	0.9219	0.0084	0.9121	0.0196	0.9083	0.0077
$\theta_{A,13}$	0.1724	0.0337	0.1195	0.0417	0.1358	0.0307	0.1645	0.0214	0.1176	0.0640

Table 9. Parameters estimated separately for UX transitions from rotations 1 to 2 and 5 to 6, July 2001 to December 2020.

		$\gamma_{1,UE}$	$\gamma_{1,UN}$	$\gamma_{1,UM}$	$\gamma_{1,UU}$	$\gamma_{2,UE}$	$\gamma_{2,UN}$	$\gamma_{2,UM}$	$\gamma_{2,UU}$
[1]	Rotation 1 estimate	0.3214	0.2145	0.0951	0.3690	0.1168	0.2364	0.0764	0.5704
[2]	Standard error	0.0061	0.0033	0.0052		0.0094	0.0089	0.0051	
[3]	Rotation 5 estimate	0.3445	0.2145	0.0908	0.3502	0.1227	0.2234	0.0717	0.5822
[4]	Standard error	0.0104	0.0043	0.0021		0.0085	0.0068	0.0058	

Table 10. Parameters estimated jointly across all rotations, July 2001 to December 2020.

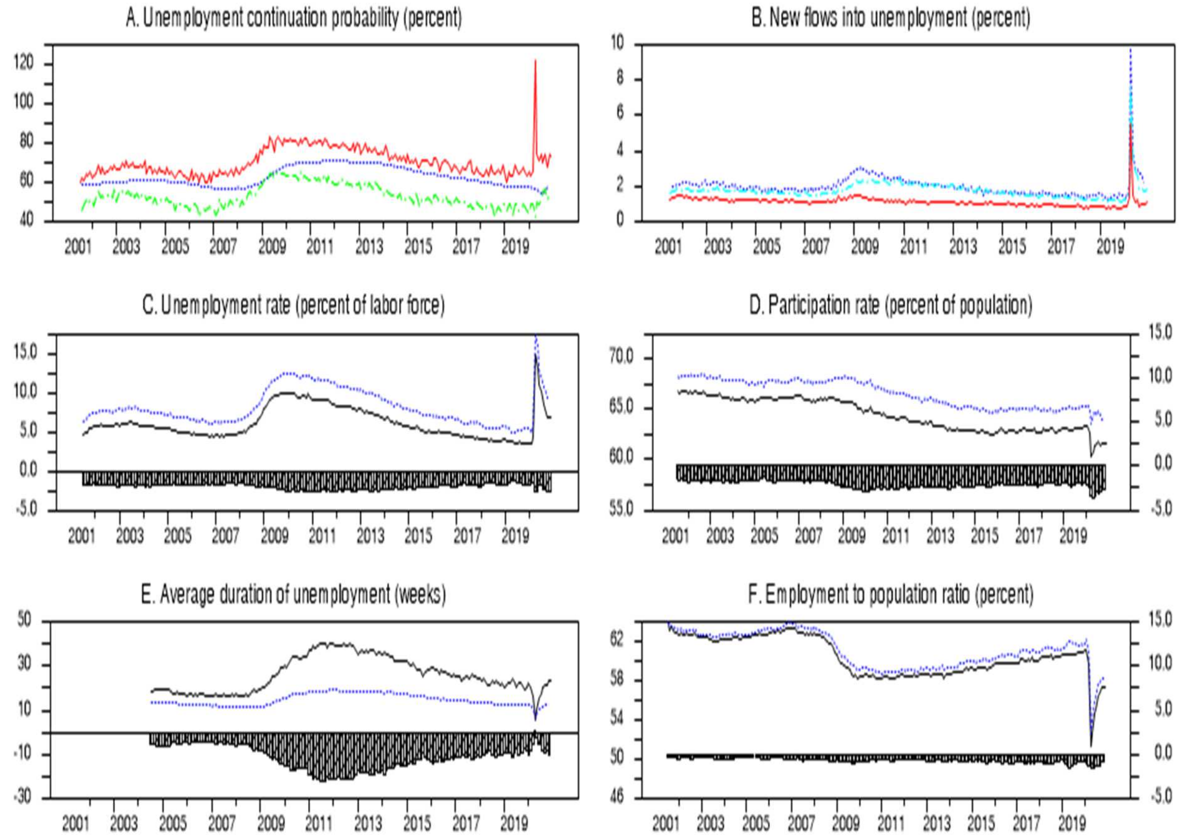
	estimate		estimate		estimate		estimate		estimate		estimate
p_1	0.8115	$\theta_{A,6}$	0.8893	$q_{1,EU}$	0.2191	$q_{1,NU}$	0.0861	$q_{1,MU}$	0.0959	$\gamma_{1,UE}$	0.3322
p_2	0.9727	$\theta_{A,7}$	0.9545	$q_{2,EU}$	0.1833	$q_{2,NU}$	0.0816	$q_{2,MU}$	0.0934	$\gamma_{1,UN}$	0.2145
w_1	0.4009	$\theta_{A,8}$	0.8013	$q_{3,EU}$	0.1964	$q_{3,NU}$	0.0764	$q_{3,MU}$	0.0970	$\gamma_{1,UM}$	0.0931
$\theta_{A,1}$	0.1434	$\theta_{A,9}$	0.9399	$q_{4,EU}$	0.1177	$q_{4,NU}$	0.0661	$q_{4,MU}$	0.0702	$\gamma_{1,UU}$	0.3602
$\theta_{A,2}$	0.7230	$\theta_{A,10}$	0.1667	$q_{5,EU}$	0.1490	$q_{5,NU}$	0.1875	$q_{5,MU}$	0.2253	$\gamma_{2,UE}$	0.1196
$\theta_{A,3}$	0.4628	$\theta_{A,11}$	0.5147	$q_{6,EU}$	0.1346	$q_{6,NU}$	0.5023	$q_{6,MU}$	0.4183	$\gamma_{2,UN}$	0.2301
$\theta_{A,4}$	0.8763	$\theta_{A,12}$	0.9048							$\gamma_{2,UM}$	0.0742
$\theta_{A,5}$	0.7026	$\theta_{A,13}$	0.1326							$\gamma_{2,UU}$	0.5761

Notes to Table 10. Also estimated (but not reported) are separate coefficients $\pi_{XE}, \pi_{XN}, \pi_{XM}, \pi_{XU}$ for $X \in \{E, N, M\}$.

Table 11. Adjusted and unadjusted estimates of duration of unemployment, July 2004 to November 2020.

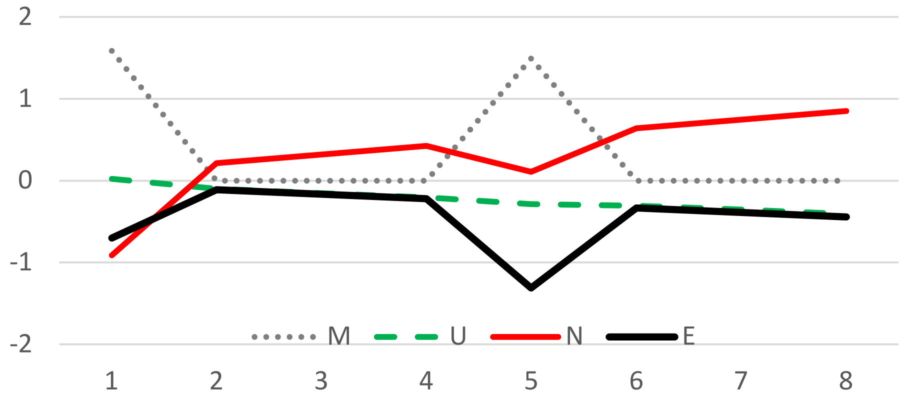
	BLS	Adjusted
< 5 weeks	29.0%	36.7%
5-14 weeks	27.9%	32.0%
15-26 weeks	15.7%	17.7%
> 26 weeks	27.5%	12.6%
Average duration	26.1 weeks	14.7 weeks

Figure 1. Alternative measures of unemployment-continuation probability, new inflows to unemployment, unemployment rate, labor force participation rate, average duration of unemployment, and employment-population ratio, July 2001 to December 2020.



Notes to Figure 1. Color key: black: BLS estimates; dotted blue: adjusted estimates; shaded bars: difference between BLS and adjusted estimates; solid red: estimates based on number of unemployed reporting duration 5 weeks or greater; dashed green: estimates based on reported transitions; dashed turquoise: BLS adjusted flows. Panel A: probability that an unemployed individual will still be unemployed next month as calculated by: (1) ratio of unemployed with duration 5 weeks or greater in month t to total unemployed in t -1 (red); (2) fraction of those unemployed in t -1 who are still unemployed in t (green); (3) reconciled estimate (blue). Panel B: Number of newly unemployed as a percent of the noninstitutional adult population as calculated by: (1) number of unemployed with duration less than 5 weeks (red); (2) *EU* and *NU* flows as adjusted by BLS (turquoise); (3) reconciled estimate (blue). Panel C: unemployment rate as calculated by BLS (black) and adjusted estimate (blue). Panel D: labor-force participation rate as calculated by BLS (black) and adjusted estimate (blue). Panel E: Average duration of unemployment as calculated by BLS (black) and adjusted estimate (blue). Panel F: Employment-to-population ratio as estimated by BLS (black) and adjusted (blue). All series seasonally adjusted.

Figure 2. Effect of rotation group on percentage of sampled individuals with indicated reported status, July 2001 to February 2020.



Notes to Figure 2. Graph shows predicted values implied by regression (3).

Figure 3. Fraction of individuals reporting labor status *E, N, M,* or *U* in each rotation group (solid blue) and fraction predicted to report that status for that rotation (dashed red), July 2001 to December 2020.

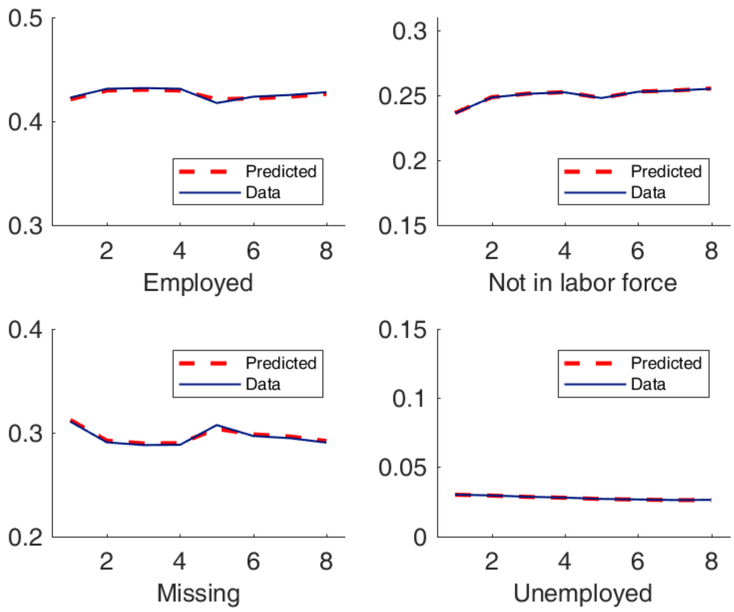


Figure 4. Actual reported transition probabilities for each rotation (solid blue) and predicted fraction (dashed red), August 2001 to December 2020.

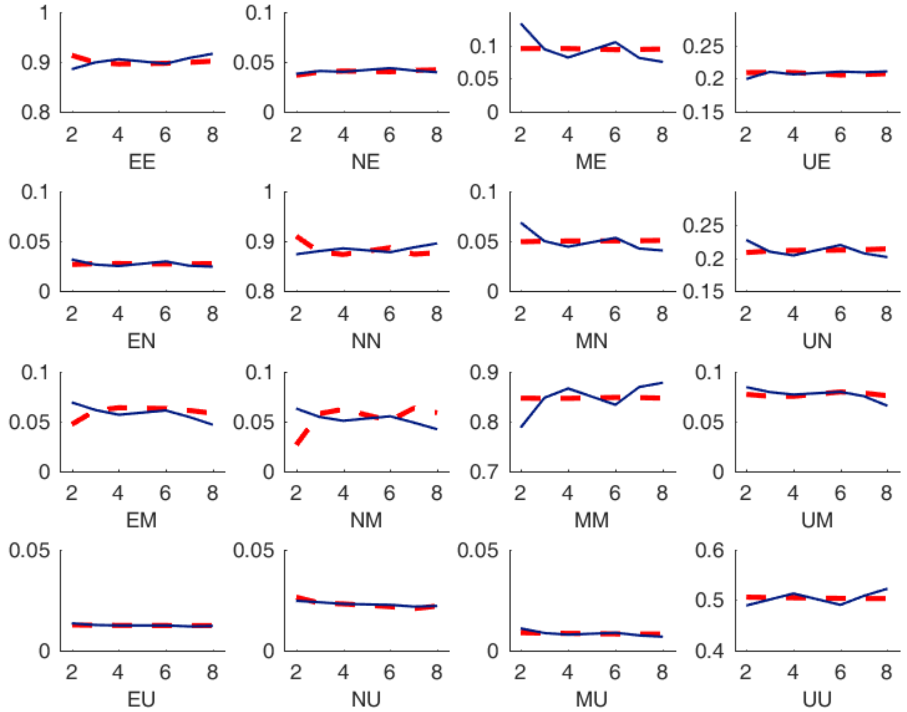


Figure 5. Changes in rotation bias parameters over time, July 2001 to December 2020.

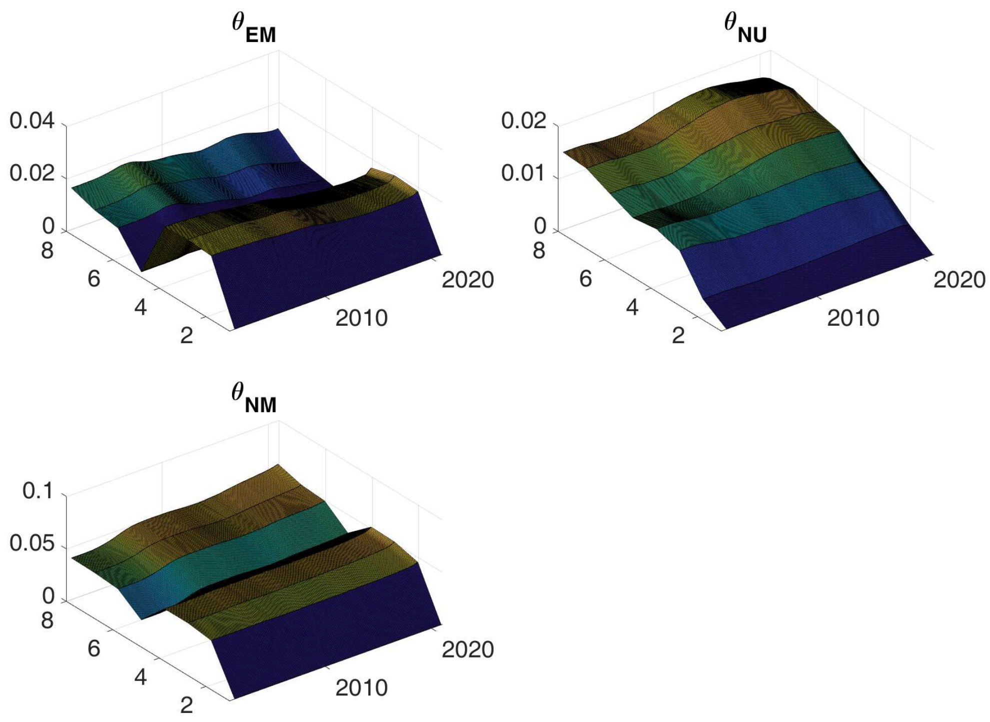
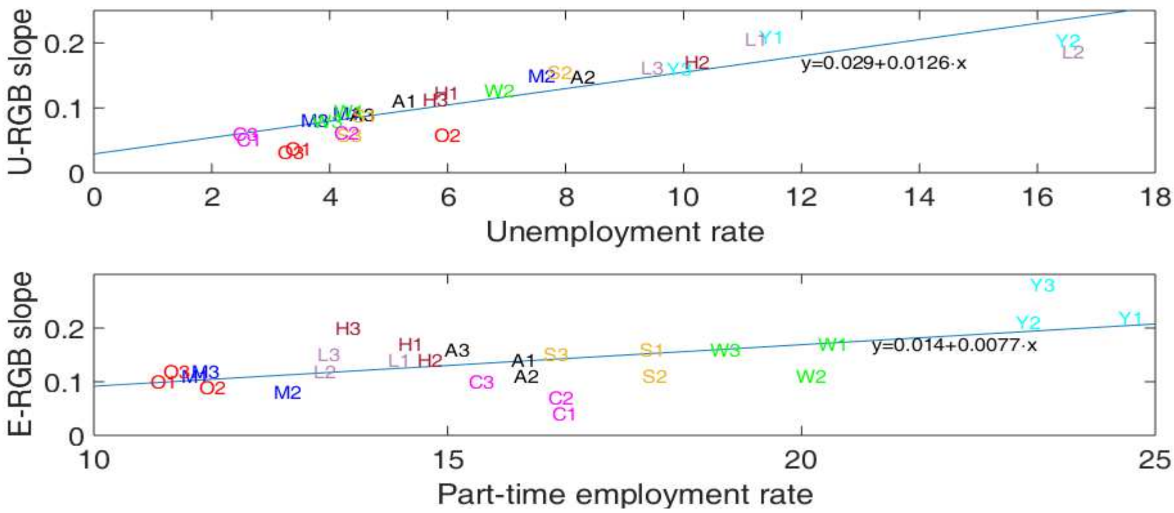


Figure 6. Measures of rotation bias, unemployment, and part-time employment for overall population and different demographic groups over three different subsamples.



Notes to Figure 6. Top panel: horizonal axis: $U/(U+E)$ as reported by BLS; vertical axis: negative of slope of regression of $U^{[j]}/(U^{j}+E^{[j]})$ on month-in-sample *j*. Demographic groups: A: Total population, Y: aged 16-24; M: men aged 25-54; W: women aged 25-54; O: aged 55 and over; L: less than high school education, H: high school graduate; S: some college; C: college graduates. Sample 1: 2001:7-2008:6; sample 2: 2008:7-2014:6; sample 3: 2014:7-2020:1. Also shown is regression line fitted to the 27 observations. Bottom panel: horizontal axis: $\dot{E}/(U+E+N)$ as reported by BLS for $\dot{E}$ part-time employment plus usual full-time workers not at work this month; vertical axis: negative of slope of regression of $\dot{E}^{[j]}/(U^{j}+E^{[j]}+N^{[j]})$ on month-in-sample *j*.

Figure 7. Contributions of different adjustments to the labor-force participation and unemployment rates, seasonally adjusted, July 2001 to December 2020.

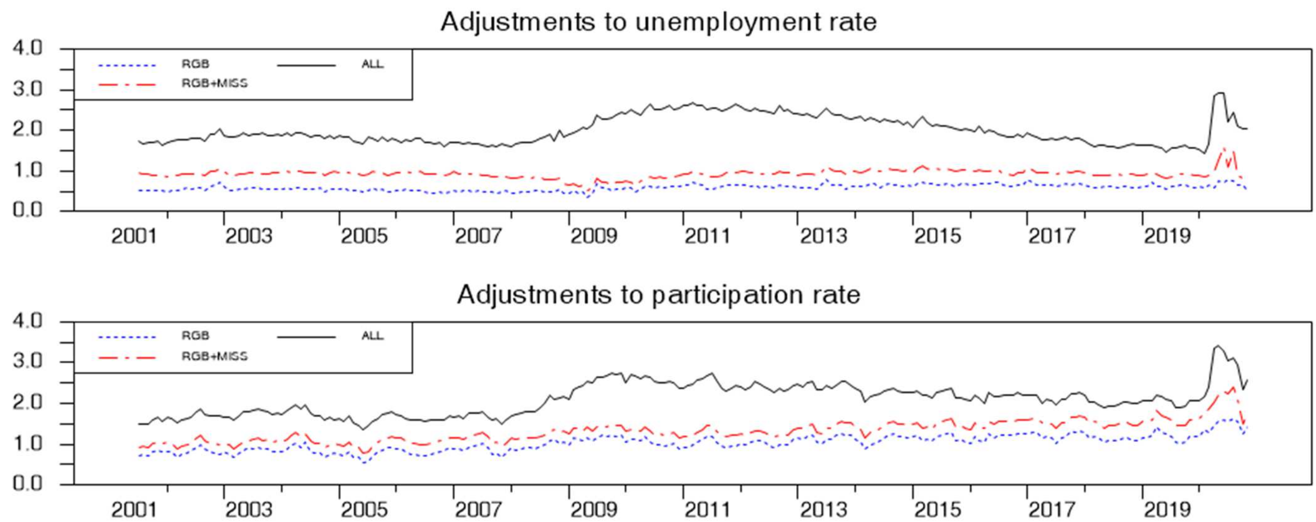


Figure 8. Characteristics of missing observations, July 2001 to November 2020.

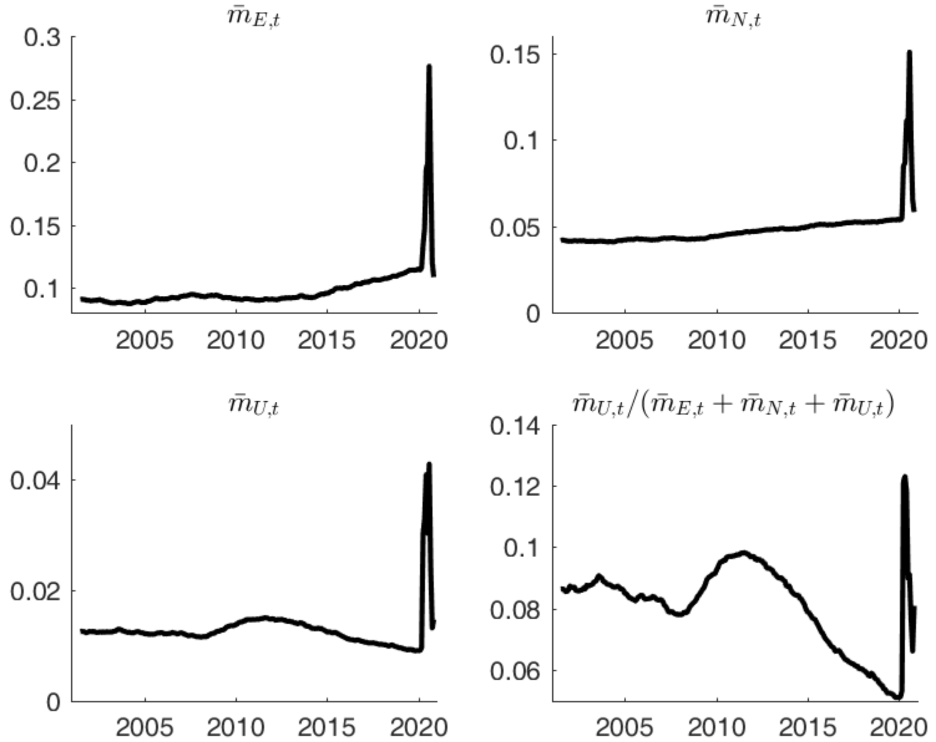
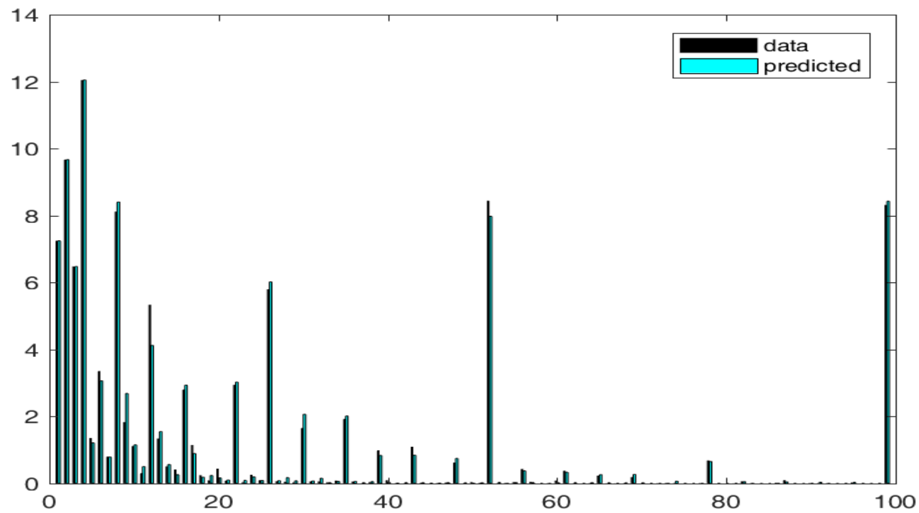
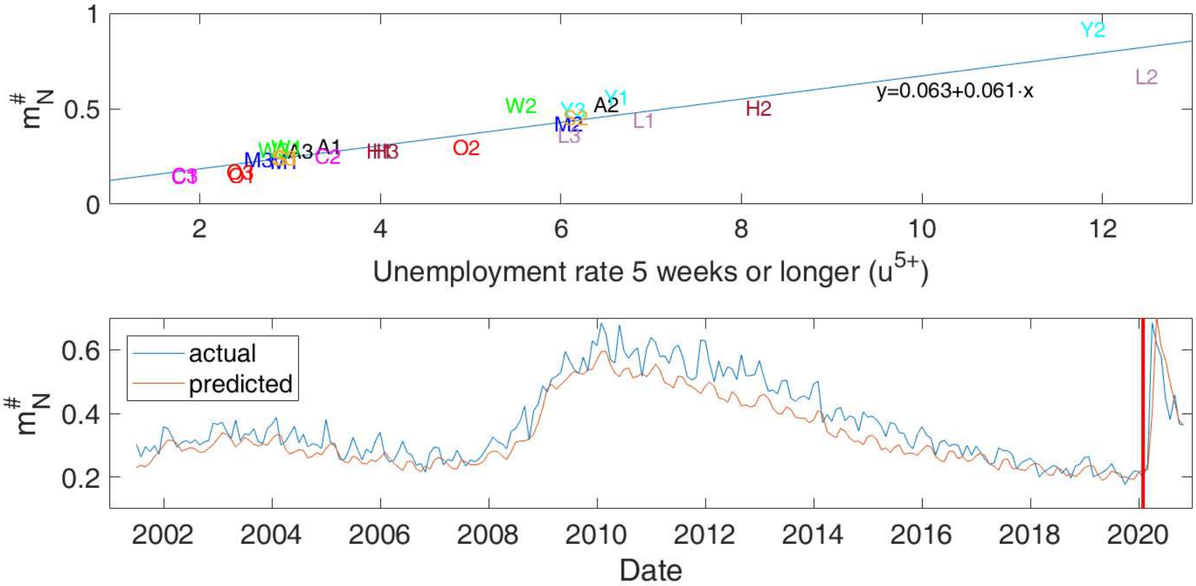


Figure 9. Predicted and reported unemployment durations in rotation 2 for individuals who were not in the labor force in rotation 1 and unemployed in rotation 2, August 2001 to December 2020.



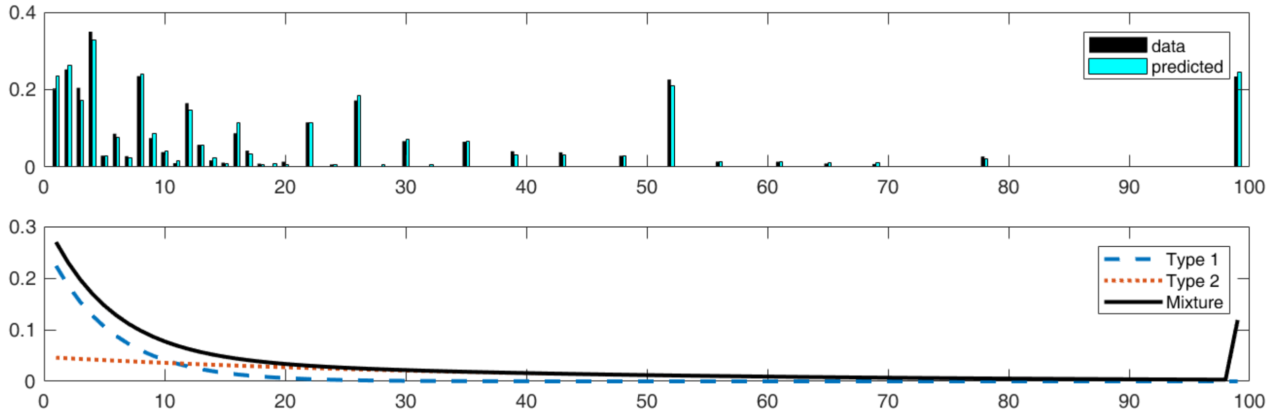
Notes to Figure 9. Horizontal axis: duration of unemployment spell in weeks. Vertical axis: of the individuals who were not in the labor force in rotation 1 and unemployed in rotation 2, the percent who reported having been searching for work at the time of rotation 2 for the indicated duration.

Figure 10. BLS reported medium- and long-term unemployment rate (u^{5+}) and percentage of individuals reporting *N* followed by U^{5+} ($m_N^{\#}$) for different demographic groups and three different samples plus predicted and actual aggregate values each month.



Notes to Figure 10. Top panel: horizonal axis: $U^{5+}/(U+E)$ as reported by BLS; vertical axis: percentage of population reporting *N* followed by U^{5+}. Demographic groups: A: Total population, Y: aged 16-24; M: men aged 25-54; W: women aged 25-54; O: aged 55 and over; L: less than high school education, H: high school graduate; S: some college; C: college graduates. Sample 1: 2001:7-2008:6; sample 2: 2008:7-2014:6; sample 3: 2014:7-2020:1. Also shown is regression line fitted to the 27 observations. Bottom panel: horizontal axis: month (2001:7 to 2020:12); vertical axis: actual percentage of population reporting *N* followed by U^{5+} and value predicted from applying the coefficients from demographic regression in the top panel to the aggregate value of $U^{5+}/(U+E)$ the previous month. Red vertical line denotes beginning of pure out-of-sample predictions.

Figure 11. Predicted and reported durations of unemployment for individuals in rotation 1, July 2001 to December 2020.



Notes to Figure 11. Top panel: reported percentage (in black) and predicted by equation (30) (in cyan) of individuals in rotation 1 who report being unemployed and having been searching for τ weeks. Bottom panel: black curve gives implied percentage of all rotation 1 individuals who are unemployed and have perceived duration of τ weeks; red and blue give percentages of each type.

Figure 12. Predicted and actual probability that someone with unemployment duration of τ weeks in rotation 1 will still be unemployed in rotation 2 (top panel) and probability $\eta_2(\tau)$ that the individual is type 2 as a function of the reported duration using parameters in Table 10 (bottom panel), August 2001 to December 2020.

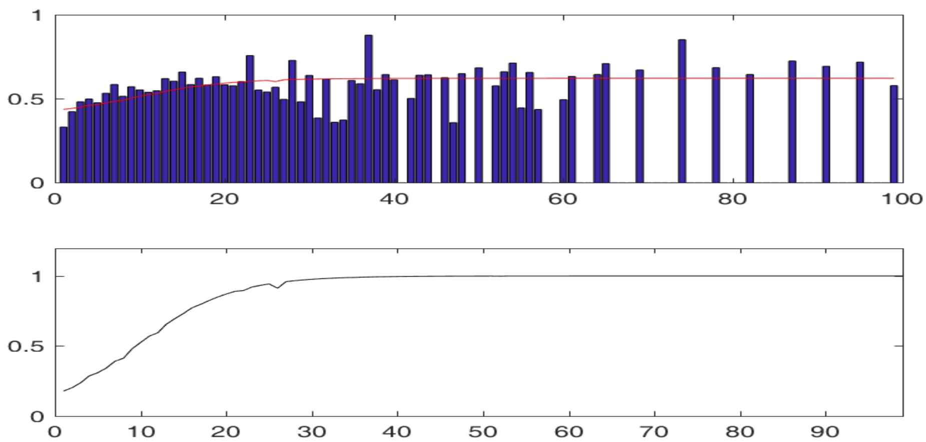
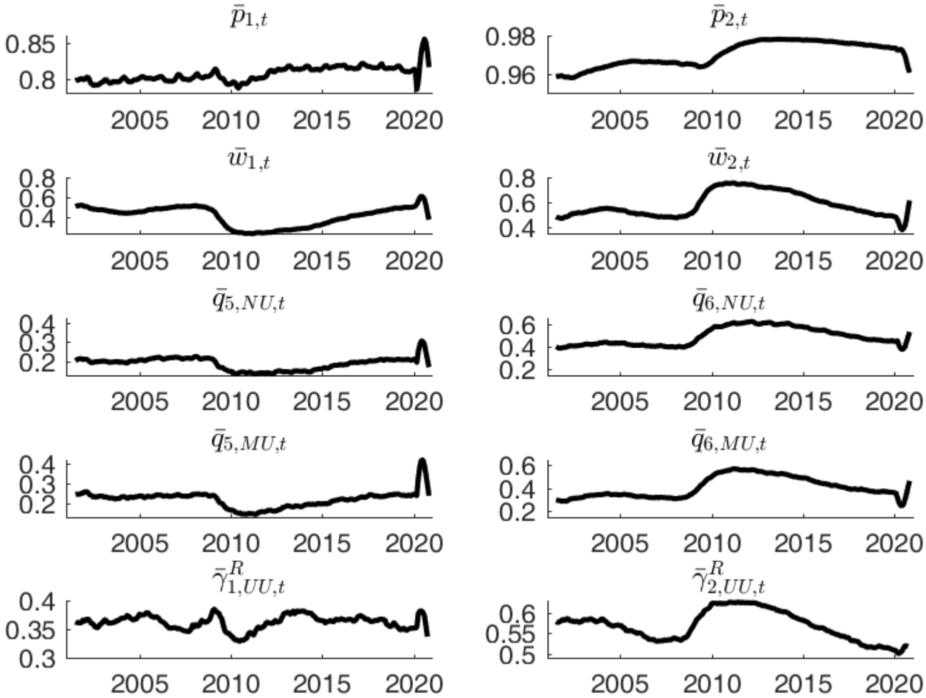


Figure 13. Time variation in selected parameters, August 2001 to December 2020.



Notes to Figure 13. Black lines denote smoothed data summaries $\bar{\theta}_t$.

Figure 14. Comparison of different adjustments, July 2001 to December 2020.