Online appendix for "Measuring Labor-Force Participation and the Incidence and Duration of Unemployment"

Hie Joo Ahn

James D. Hamilton

March 6, 2021

Appendix A. Details of data construction.

Calculating weights for each individual. For any individual i we start with the weight w_i assigned to that individual by the BLS for the first month that a weight was reported for that individual.[38] Note this is a function of the individual and not a function of time. For the month when a weight was first reported for that individual, we calculated the average value of w_i for all individuals who were either E, N, M, or U in that month. Call this number $\bar{w}_i$. Again, $\bar{w}_i$ is a fixed number for any given individual. To rescale this to an average value around 1 we assign the weight $\omega_i = w_i/\bar{w}_i$ to individual i.

Our raw data (denoted $\tilde{y}$) thus consist of weighted counts of individuals in various categories. Let $\tilde{y}^{[j]}_{X,t}$ be the weighted count of the number of individuals with status $X \in \{E, N, M, U\}$ when they would have been interviewed in rotation j in month t before making any adjustments for the start and end of the sample or sample redesigns and $y^{[j]}_{X,t}$ the value after making the adjustments below. Likewise, let $\tilde{y}^{[j]}_{X_1,X_2,t}$ be the unadjusted weighted count of the number with status X_1 in rotation $j-1$ in $t-1$ and status X_2 in rotation j in t and $y^{[j]}_{X_1,X_2,t}$ the adjusted estimate.

Adjusting missing observations at the start of the sample. If the sample begins in $t_0 =$ 2001:7, we will miss individuals who would have had a history like $EMMM-MMMM$ if their first interview had been at t_0-1; such an individual is tracked in the column of Table A-2 associated with MIS in 2001:7 equal to 2, but would never appear in the data set. This will cause $\tilde{y}^{[2]}_{M,t_0}$ to be lower than subsequent observations that would have complete histories on individuals like this, such as the MIS in 2001:7 = 1 column. We calculate the average value of $\tilde{y}^{[2]}_{M,t}$ for the first year of complete observations on this magnitude to form an estimate of the amount by which $\tilde{y}^{[2]}_{M,t_0}$ is an underestimate of the true value:

$$\Delta^{[2]}_{t_0} = (1/12)\sum_{t=t_0+1}^{t_0+12} \tilde{y}^{[2]}_{M,t} - \tilde{y}^{[2]}_{M,t_0}.$$

[38] In data since 1998 this is the variable pwcmpwgt.

We accordingly add $\Delta_{t_0}^{[2]}$ to the estimated count of people who transitioned from M in rotation 2 in t_0 to M in rotation 3 in t_0+1:

$$y_{M,M,t_0+1}^{[3]} = \tilde{y}_{M,M,t_0+1}^{[3]} + \Delta_{t_0}^{[2]}. \tag{A1}$$

Looking at the 2001:8 observation in the MIS in 2001:7 = 2 column of Table A-2, we should also adjust the count of people who are imputed to have made a transition from M in rotation 3 in t_0+1 to M in rotation 4 in t_0+2:

$$y_{M,M,t_0+2}^{[4]} = \tilde{y}_{M,M,t_0+2}^{[4]} + \Delta_{t_0}^{[2]}. \tag{A2}$$

Notice that since the unadjusted data $\tilde{y}$ satisfy the internal consistency conditions (1) and (2),

$$\begin{aligned}
\tilde{y}_{M,t_0+1}^{[3]} &= \tilde{y}_{E,M,t_0+1}^{[3]} + \tilde{y}_{N,M,t_0+1}^{[3]} + \tilde{y}_{M,M,t_0+1}^{[3]} + \tilde{y}_{U,M,t_0+1}^{[3]} \\
&= \tilde{y}_{M,E,t_0+2}^{[4]} + \tilde{y}_{M,N,t_0+2}^{[4]} + \tilde{y}_{M,M,t_0+2}^{[4]} + \tilde{y}_{M,U,t_0+2}^{[4]},
\end{aligned}$$

so do the adjusted data with $y_{M\,t_0+1}^{[3]} = \tilde{y}_{M,t_0+1}^{[3]} + \Delta_{t_0}^{[2]}$:

$$\begin{aligned}
y_{M,t_0+1}^{[3]} &= y_{E,M,t_0+1}^{[3]} + y_{N,M,t_0+1}^{[3]} + y_{M,M,t_0+1}^{[3]} + y_{U,M,t_0+1}^{[3]} \\
&= \tilde{y}_{E,M,t_0+1}^{[3]} + \tilde{y}_{N,M,t_0+1}^{[3]} + \tilde{y}_{M,M,t_0+1}^{[3]} + \Delta_{t_0}^{[2]} + \tilde{y}_{U,M,t_0+1}^{[3]} \\
&= \tilde{y}_{M,t_0+1}^{[3]} + \Delta_{t_0}^{[2]} \\
&= y_{M,E,t_0+2}^{[4]} + y_{M,N,t_0+2}^{[4]} + y_{M,M,t_0+2}^{[4]} + y_{M,U,t_0+2}^{[4]} \\
&= \tilde{y}_{M,E,t_0+2}^{[4]} + \tilde{y}_{M,N,t_0+2}^{[4]} + \tilde{y}_{M,M,t_0+2}^{[4]} + \Delta_{t_0}^{[2]} + \tilde{y}_{M,U,t_0+2}^{[4]} \\
&= \tilde{y}_{M,t_0+1}^{[3]} + \Delta_{t_0}^{[2]}.
\end{aligned}$$

Continuing down the MIS in 2001:7 = 2 column of Table A-2, we also make adjustments to the later transitions:

$$y_{M,M,t_0+12}^{[6]} = \tilde{y}_{M,M,t_0+12}^{[6]} + \Delta_{t_0}^{[2]}$$

$$y_{M,M,t_0+13}^{[7]} = \tilde{y}_{M,M,t_0+13}^{[7]} + \Delta_{t_0}^{[2]}$$

$$y^{[8]}_{M,M,t_0+14} = \tilde{y}^{[8]}_{M,M,t_0+14} + \Delta^{[2]}_{t_0}.$$

Consider next an individual who would have had a history like $EEMM - MMMM$ if they had been first interviewed in rotation 1 in $t_0 - 2$ (column MIS in 2007:1 = 3 of Table A-2). The analogous adjustments here are

$$\Delta^{[3]}_{t_0} = (1/12)\textstyle\sum_{t=t_0+2}^{t_0+13} \tilde{y}^{[3]}_{M,t} - \tilde{y}^{[3]}_{M,t_0}$$

$$y^{[4]}_{M,M,t_0} = \tilde{y}^{[4]}_{M,M,t_0} + \Delta^{[3]}_{t_0}$$

$$y^{[6]}_{M,M,t_0+11} = \tilde{y}^{[6]}_{M,M,t_0+11} + \Delta^{[3]}_{t_0}$$

$$y^{[7]}_{M,M,t_0+12} = \tilde{y}^{[7]}_{M,M,t_0+12} + \Delta^{[3]}_{t_0}$$

$$y^{[8]}_{M,M,t_0+13} = \tilde{y}^{[8]}_{M,M,t_0+13} + \Delta^{[3]}_{t_0}.$$

We also estimate the average number of M in rotation 5 over $t_0 + 15$ to $t_0 + 26$ and adjust the appropriate cohorts as follows for $j = 0, 1, ..., 9$:

$$\Delta^{[5,j]}_{t_0} = (1/12)\textstyle\sum_{t=t_0+15}^{t_0+26} \tilde{y}^{[5]}_{M,t} - \tilde{y}^{[5]}_{M,t_0+j}$$

$$y^{[6]}_{M,M,t_0+j+1} = \tilde{y}^{[6]}_{M,M,t_0+j+1} + \Delta^{[5,j]}_{t_0}$$

$$y^{[7]}_{M,M,t_0+j+2} = \tilde{y}^{[7]}_{M,M,t_0+j+2} + \Delta^{[5,j]}_{t_0}$$

$$y^{[8]}_{M,M,t_0+j+3} = \tilde{y}^{[8]}_{M,M,t_0+j+3} + \Delta^{[5,j]}_{t_0}.$$

The final adjustments are

$$\Delta^{[6]}_{t_0} = (1/12)\textstyle\sum_{t=t_0+16}^{t_0+27} \tilde{y}^{[6]}_{M,t} - \tilde{y}^{[6]}_{M,t_0}$$

$$y^{[7]}_{M,M,t_0+1} = \tilde{y}^{[7]}_{M,M,t_0+1} + \Delta^{[6]}_{t_0}$$

$$y^{[8]}_{M,M,t_0+2} = \tilde{y}^{[8]}_{M,M,t_0+2} + \Delta^{[6]}_{t_0}$$

$$\Delta^{[7]}_{t_0} = (1/12)\textstyle\sum_{t=t_0+17}^{t_0+28} \tilde{y}^{[7]}_{M,t} - \tilde{y}^{[7]}_{M,t_0}$$

$$y^{[8]}_{M,M,t_0+1} = \tilde{y}^{[8]}_{M,M,t_0+1} + \Delta^{[7]}_{t_0}.$$

On the last equation, note that $\tilde{y}^{[8]}_{M,M,t_0+1} = 0$ by construction.

Adjusting missing observations at the end of the sample. Table A-3 tracks cohorts in terms of their status at the end of the sample. The example in Table A-3 uses $T = 2020{:}2$, which was the end of the original sample used, though a later version of the paper performed exactly the same calculations using instead $T = 2020{:}12$. At the end of the sample (date T), undercounted M arise from individuals who would have later had a status like E if interviewed for the full 8 rotations after the sample end. Here the adjustments are

$$\Delta^{[1,j]}_T = (1/12)\textstyle\sum_{t=T-15}^{T-26} \tilde{y}^{[1]}_{M,t} - \tilde{y}^{[1]}_{M,T-j} \quad \text{for } j = 0, 1, ..., 14$$

$$y^{[1]}_{M,T} = \tilde{y}^{[1]}_{M,T} + \Delta^{[1,0]}_T$$

$$y^{[2]}_{M,M,T-j} = \tilde{y}^{[2]}_{M,M,T-j} + \Delta^{[1,j+1]}_T \quad \text{for } j = 0, 1, ..., 13$$

$$y^{[3]}_{M,M,T-j} = \tilde{y}^{[3]}_{M,M,T-j} + \Delta^{[1,j+2]}_T \quad \text{for } j = 0, 1, ..., 12$$

$$y^{[4]}_{M,M,T-j} = \tilde{y}^{[4]}_{M,M,T-j} + \Delta^{[1,j+3]}_T \quad \text{for } j = 0, 1, ..., 11$$

$$y^{[5]}_{M,T} = \tilde{y}^{[5]}_{M,T} + \Delta^{[1,12]}_T$$

$$y^{[6]}_{M,M,T-j} = \tilde{y}^{[6]}_{M,M,T-j} + \Delta^{[1,j+13]}_T \quad \text{for } j = 0, 1$$

$$y^{[7]}_{M,M,T} = \tilde{y}^{[7]}_{M,M,T} + \Delta^{[1,14]}_T.$$

Adjusting missing observations from the sample redesign of CPS in 2004 and 2014. At date $t_2 = 2004{:}8$ the BLS dropped some households who otherwise would have been included in rotation 5 and added some new households in order to better represent the U.S. population, and did the same thing to the rotation 5 individuals for each of the following 11 months. Individuals added into rotation 5 create a bulge in histories like $MMMM - EEEE$ for people who would have been in rotation 1 over $t_2 - 12$ through $t_2 - 1$ (see Table A-4). We adjust for this by comparing the number of M in rotation 1 during this period with the average value over the 12 months beginning with t_2:

$$\Delta^{[1,j]}_{t_2} = \tilde{y}^{[1]}_{M,t_2-j} - (1/12)\textstyle\sum_{t=t_2}^{t_2+11} \tilde{y}^{[1]}_{M,t} \quad \text{for } j = 1, ..., 12.$$

We then reduce the estimates of MM transitions for $j = 1$ to 12 accordingly:

$$y^{[2]}_{M,M,t_2-j+1} = \tilde{y}^{[2]}_{M,M,t_2-j+1} - \Delta^{[1,j]}_{t_2}$$

$$y^{[3]}_{M,M,t_2-j+2} = \tilde{y}^{[3]}_{M,M,t_2-j+2} - \Delta^{[1,j]}_{t_2}$$

$$y^{[4]}_{M,M,t_2-j+3} = \tilde{y}^{[4]}_{M,M,t_2-j+3} - \Delta^{[1,j]}_{t_2}.$$

Individuals dropped from rotation 5 at $t_2, t_2+1, ..., t_2+11$ create artificial histories like $EEEE-MMMM$. We adjust for these as follows for $j = 0, 1, ..., 11$:

$$\Delta^{[5,j]}_{t_2} = \tilde{y}^{[5]}_{M,t_2+j} - (1/12) \sum_{t=t_2-2}^{t_2-13} \tilde{y}^{[5]}_{M,t}$$

$$y^{[6]}_{M,M,t_2+j+1} = \tilde{y}^{[6]}_{M,M,t_2+j+1} - \Delta^{[5,j]}_{t_2}$$

$$y^{[7]}_{M,M,t_2+j+2} = \tilde{y}^{[7]}_{M,M,t_2+j+2} - \Delta^{[5,j]}_{t_2}$$

$$y^{[8]}_{M,M,t_2+j+3} = \tilde{y}^{[8]}_{M,M,t_2+j+3} - \Delta^{[5,j]}_{t_2}.$$

There was a similar change in population controls at date $t_3 = 2014{:}8$, dealt with using the identical adjustments replacing t_2 with t_3.

Effects of adjustments. Figure A-1 shows the effects of these adjustments. The red dashed lines plot transitions $\tilde{y}^{[j]}_{M,M,t}$ before adjustments, exhibiting bulges prior to the 2004 and 2014 sample redesign and dips at the start and end of the sample. The solid blue lines plot $y^{[j]}_{M,M,t}$ which are the series used in our analysis.

Appendix B. Modeling number preference.

Even number bias. For durations of 8 weeks or fewer, people are more likely to report an even number than an odd number. One way to represent this is to suppose that people whose true or perceived duration is 1, 3, 5, or 7 weeks have a probability $(1 - \theta_{A,1})$ of reporting their duration correctly and a probability $\theta_{A,1}$ of instead reporting 2, 4, 6, or 8 weeks. If this was the only source of reporting error, the matrix A would take the following form:[39]

$$\underset{(99\times 99)}{A_1} = \begin{bmatrix} \Lambda_1 & 0_{8,91} \\ 0_{91,8} & I_{91} \end{bmatrix}$$

$$\underset{(8\times 8)}{\Lambda_1} = \begin{bmatrix} \Theta_1 & 0_{2,2} & 0_{2,2} & 0_{2,2} \\ 0_{2,2} & \Theta_1 & 0_{2,2} & 0_{2,2} \\ 0_{2,2} & 0_{2,2} & \Theta_1 & 0_{2,2} \\ 0_{2,2} & 0_{2,2} & 0_{2,2} & \Theta_1 \end{bmatrix}$$

$$\underset{(2\times 2)}{\Theta_1} = \begin{bmatrix} 1-\theta_{A,1} & 0 \\ \theta_{A,1} & 1 \end{bmatrix}.$$

For example, if the true duration $\tau = 3$ (third column of A_1) there would be a probability $1 - \theta_{A,1}$ of reporting $\tau = 3$ represented by the (3,3) element of A_1 and a probability $\theta_{A,1}$ of reporting $\tau = 4$.

Reporting as nearest month. After the 1994 redesign, respondents were given the option of reporting unemployment duration in months or years rather than weeks ("flexible reporting periodicity"), although they are explicitly asked for duration in weeks if they report four or fewer months of unemployment. Even if someone has been unemployed for fewer than four months during the reference week, the person is likely to report their duration in integer months rather than the actual number of weeks. This shows up in an unusual concentration of reported durations at 4 weeks, 8-9 weeks, 12-13 weeks, and 16-17 weeks. We represent this by allowing that a fraction $\theta_{A,2}$ of those who are truly unemployed for 5 weeks report the duration as 4, and the same fraction $\theta_{A,2}$ of those who are unemployed for 7, 11, or 15 weeks report their durations as 8, 12, or 16. We suppose that of those whose true duration is 6, 10, or 14, a fraction $\theta_{A,3}/2$ round the month down (to 4, 8, or 12) and another $\theta_{A,3}/2$ round up (to 8, 12, or 16). We don't find evidence in the data that durations 9, 13, or 17 get particularly rounded down. We capture these tendencies in a

[39] Here $0_{n,m}$ denotes an $(n \times m)$ matrix of zeros and I_n the $(n \times n)$ identity matrix.

(99×99) matrix A_2 whose rows 4-12 and columns 4-11 are given by

$$\underset{(9\times 8)}{\Lambda_2} = \begin{array}{c} 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{array} \begin{bmatrix} 1 & \theta_{A,2} & \theta_{A,3}/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\theta_{A,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\theta_{A,3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\theta_{A,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{A,3}/2 & \theta_{A,2} & 1 & 0 & \theta_{A,3}/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-\theta_{A,3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\theta_{A,2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{A,3}/2 & \theta_{A,2} \end{bmatrix}.$$

Rows 12-17 and columns 12-17 of A_2 are

$$\underset{(6\times 6)}{\Lambda_3} = \begin{array}{c} 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \end{array} \begin{bmatrix} 1 & 0 & \theta_{A,3}/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\theta_{A,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\theta_{A,2} & 0 & 0 \\ 0 & 0 & \theta_{A,3}/2 & \theta_{A,2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For unemployment spells longer than 4 months the rounding becomes even more pronounced, as the CPS allows respondents to report an interval in months and years. Census Bureau multiply a conversion factor of 4.33 to the reported duration in months and 52 to that in years, which then gets recorded as $\tau \in \Omega = \{16, 22, 26, 30, 35, 39, 43, 48, 52, 56, 61, 65, 69, 74, 78, 82, 87, 91, 95\}$ weeks. We represent the increased propensity to round as durations get longer by replacing $\theta_{A,2}$ and $\theta_{A,3}$, which governed the probabilities that a duration got rounded one week or two weeks to the nearest month, with analogous (but larger) probabilities $\theta_{A,4}$ and $\theta_{A,5}$ for durations between 18 and 32 weeks, even larger probabilities $\theta_{A,6}$ and $\theta_{A,7}$ for durations between 33 and 52 weeks, and $\theta_{A,8}$ and $\theta_{A,9}$ for durations 53 weeks and longer. Specifically, if someone is unemployed between 18 and 32 weeks and is exactly 1 week away from an element of Ω, there is a probability $\theta_{A,4}$ that the duration gets reported as the nearest month. That is, for durations between 18 and 32 weeks, rows $\tau - 3$ through $\tau + 3$ of column τ of A_2 are given by Λ_4 for $\tau \in \Omega$, Λ_5 for durations one week less than an element of Ω (namely $\tau \in \{21, 25, 29\}$), and Λ_6 for durations one week more than an element of Ω (namely $\tau \in \{23, 27, 31\}$), where

$$\Lambda_4 = \begin{array}{c} \tau-3 \\ \tau-2 \\ \tau-1 \\ \tau \\ \tau+1 \\ \tau+2 \\ \tau+3 \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Lambda_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1-\theta_{A,4} \\ \theta_{A,4} \\ 0 \\ 0 \end{bmatrix} \quad \Lambda_6 = \begin{bmatrix} 0 \\ 0 \\ \theta_{A,4} \\ 1-\theta_{A,4} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

If the person is exactly two weeks or exactly three weeks away from two different elements of Ω, with probability $\theta_{A,5}/2$ it gets rounded up and probability $\theta_{A,5}/2$ it gets rounded down. If the person is two weeks away from only a single element of Ω, we assume that it gets rounded to the nearest month with probability $(\theta_{A,4} + \theta_{A,5})/2$. Thus for durations between 18 and 32 weeks and exactly two weeks away from two elements of Ω (namely $\tau \in \{24, 28\}$), rows $\tau - 3$ through $\tau + 3$ of column τ of A_2 are given by Λ_7 below. The vector Λ_8 is used for durations two weeks below a single element of Ω ($\tau = 20$), Λ_9 for elements two weeks above a single element of Ω (namely $\tau = 18$), and Λ_{10} for durations three weeks away from two elements of Ω ($\tau = 19$):

$$\Lambda_7 = \begin{array}{c} \tau-3 \\ \tau-2 \\ \tau-1 \\ \tau \\ \tau+1 \\ \tau+2 \\ \tau+3 \end{array} \begin{bmatrix} 0 \\ \theta_{A,5}/2 \\ 0 \\ 1-\theta_{A,5} \\ 0 \\ \theta_{A,5}/2 \\ 0 \end{bmatrix} \quad \Lambda_8 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1-\frac{\theta_{A,4}+\theta_{A,5}}{2} \\ 0 \\ \frac{\theta_{A,4}+\theta_{A,5}}{2} \\ 0 \end{bmatrix} \quad \Lambda_9 = \begin{bmatrix} 0 \\ \frac{\theta_{A,4}+\theta_{A,5}}{2} \\ 0 \\ 1-\frac{\theta_{A,4}+\theta_{A,5}}{2} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Lambda_{10} = \begin{bmatrix} \theta_{A,5}/2 \\ 0 \\ 0 \\ 1-\theta_{A,5} \\ 0 \\ 0 \\ \theta_{A,5}/2 \end{bmatrix}.$$

For $\tau = 32$ we use Λ_{13} below with $\theta_{A,7}$ replaced by $\theta_{A,5}$.

For durations between 33 weeks and 52 weeks, we use analogous matrices with $\theta_{A,4}$ replaced by $\theta_{A,6}$ and $\theta_{A,5}$ replaced by $\theta_{A,7}$. Specifically, for observations exactly one week below or one week above an element of Ω we use Λ_5 or Λ_6 with $\theta_{A,4}$ replaced by $\theta_{A,6}$. For observations exactly two weeks from two different elements of Ω (namely $\tau \in \{37, 41, 50\}$) we use Λ_{11}, for observations exactly two weeks below a single element of Ω (namely $\tau = 46$) we use Λ_{12}, and for observations exactly two weeks above a single element of Ω (namely $\tau = 45$) we use Λ_{13}:

$$\Lambda_{11} = \begin{array}{c} \tau-3 \\ \tau-2 \\ \tau-1 \\ \tau \\ \tau+1 \\ \tau+2 \\ \tau+3 \end{array} \begin{bmatrix} 0 \\ \theta_{A,7}/2 \\ 0 \\ 1-\theta_{A,7} \\ 0 \\ \theta_{A,7}/2 \\ 0 \end{bmatrix} \qquad \Lambda_{12} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1-\theta_{A,7} \\ 0 \\ \theta_{A,7} \\ 0 \end{bmatrix} \qquad \Lambda_{13} = \begin{bmatrix} 0 \\ \theta_{A,7} \\ 0 \\ 1-\theta_{A,7} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

For durations between 53 and 99 weeks we replace $\theta_{A,6}$ and $\theta_{A,7}$ by $\theta_{A,8}$ and $\theta_{A,9}$, respectively. We set the upper-left block of A_2 to I_3 and all other elements to 0. The model of measurement error so far is then summarized by the product $A_2 A_1$.

6-month interval reporting. Finally there is a tendency to report longer durations as either 6 months (26 weeks), 12 months (52 weeks), 18 months (78 weeks), or longer than 99 weeks (top code). To capture this, we adapt the heaping model of Torelli and Trivellato (1993). We assume that individuals may begin to report their duration as the closest multiple of 6 months after 3 months of unemployment. Define a (99×99) matrix A_3 whose upper-left block is given by I_{13}. For every individual who from the preceding specifications would have reported an unemployment spell between 14 and 38 weeks (or between 3 and 7 months), there is a probability $\theta_{A,10}$ that they don't report τ but instead report 6 months. In other words, for $\tau \in \{14, 15, ..., 25, 27, 28, ..., 38\}$ the row τ, column τ element of A_3 is given by $1 - \theta_{A,10}$ and the row 26, column τ element is $\theta_{A,10}$. Likewise, if someone has been unemployed for 39 to 51 weeks, the person may report 12 months with probability $\theta_{A,11}$ (row τ, column τ of A_2 is $1 - \theta_{A,11}$ and row 52, column τ is $\theta_{A,9}$). If someone has been unemployed 12-15 months, we assume that the person would report 12 months with probability $\theta_{A,11}$ or possibly instead report the top-code duration (longer than 99 weeks) with probability $\theta_{A,13}$. That is, for $\tau \in \{53, 54, ..., 72\}$, column τ of A_3 has $1 - \theta_{A,11} - \theta_{A,13}$ in the τth row, $\theta_{A,11}$ in row 52, and $\theta_{A,13}$ in the last row.

For intervals longer than 15 months, there is a probability $\theta_{A,12}$ of reporting the top code, so that for $\tau \in \{73, 74, 75, 76, 77, 79, 80, 81, ..., 98\}$ we set the τth row of column τ of A_3 to $1 - \theta_{A,12}$, the 99th row to $\theta_{A,12}$ and all other elements to zero. The reason we leave a 78-week reporting interval out of this set is that we observe a modest spike in reporting intervals of 78 weeks (18 months), and so take the 78th column of A_3 to be the 78th column of I_{99}. The last column of A_3 is of course the last column of I_{99}.

Our model of measurement error is thus

$$\underset{(99\times99)}{A} = \underset{(99\times99)}{A_3}\underset{(99\times99)}{A_2}\underset{(99\times99)}{A_1}.$$

The value of A is a function of $\theta_A = (\theta_{A,1}, \theta_{A,2}, ..., \theta_{A,13})'$ and we estimate the elements of θ_A by quasi maximum likelihood. Note that perfect reporting is allowed as a special case of this framework when $\theta_A = 0$.

Appendix C. Quasi-maximum likelihood estimation.

Here we describe how we calculated standard errors. Let n_t denote the number of individuals sampled in month t (including those categorized as M in month t) and let $n = \sum_{t=1}^{T} n_t$ be the total number of observations. For q some chosen Newey-West bandwidth (our empirical estimates use $q = 96$) define

$$\begin{aligned} h_t(\lambda_X) &= y_{E,t}\frac{\partial \ln \pi_E}{\partial \lambda_X} + y_{N,t}\frac{\partial \ln \pi_N}{\partial \lambda_X} + y_{M,t}\frac{\partial \ln \pi_M}{\partial \lambda_X} \\ &\quad + \sum_{\tau=1}^{99} y_{U,t}(\tau)\frac{\partial \ln \dot{\pi}_U(\tau)}{\partial \lambda_X} \end{aligned}$$

$$\hat{D} = -n^{-1}\sum_{t=1}^{T} \left.\frac{\partial h_t(\lambda_X)}{\partial \lambda_X'}\right|_{\lambda_X = \hat{\lambda}_X}$$

$$\hat{\Gamma}_v = n^{-1}\sum_{t=v+1}^{T} h_t(\hat{\lambda}_X) h_{t-v}(\hat{\lambda}_X)'$$

$$\hat{S} = \hat{\Gamma}_0 + \sum_{v=1}^{q}\left[1 - \frac{v}{q+1}\right](\hat{\Gamma}_v + \hat{\Gamma}_v')$$

$$\hat{V} = (\hat{D}\hat{S}^{-1}\hat{D})^{-1}. \tag{A3}$$

The square root of the (i,i) element of $n^{-1}\hat{V}$ was used to calculate a standard error for the ith element of $\hat{\lambda}_X$.

Note that it's not strictly necessary to calculate n, since it cancels out in calculation of $n^{-1}(\hat{D}\hat{S}^{-1}\hat{D})^{-1}$. We write expressions in this form because $\hat{D}$, $\hat{S}$ and $\hat{V}$ as written are consistent estimates of nondegenerate population analogues, and matrix inversions may be better behaved numerically when the expressions are calculated as written.

One can see why this works with a simple illustrative example. Let $y_{it} = 1$ if person i is employed in month t and 0 otherwise. Suppose that the probability that an individual is employed in month t is given by λ_t, so that conditional on λ_t, the mean and variance of y_{it} are $E(y_{it}|\lambda_t) = \lambda_t$ and $E[(y_{it} - \lambda_t)^2|\lambda_t] = \lambda_t(1-\lambda_t)$. Suppose that λ_t is distributed across months from some process whose mean is λ and whose vth autocovariance is γ_v ($E(\lambda_t) = \lambda$ and $E(\lambda_t - \lambda)(\lambda_{t-v} - \lambda) = \gamma_v$ for $v = 0, 1, 2, ...$). Then $y_t = \sum_{i=1}^{n_t} y_{it}$ has conditional mean $E(y_t|\lambda_t) = n_t\lambda_t$, unconditional mean $E(y_t) = n_t\lambda$, and unconditional variance $E(y_t - n_t\lambda)^2$. To evaluate the last magnitude we first take expectations conditional on λ_t,

$$\begin{aligned}
E[(y_t - n_t\lambda)^2|\lambda_t] &= E[(y_t - n_t\lambda_t + n_t\lambda_t - n_t\lambda)^2|\lambda_t] \\
&= E[(y_t - n_t\lambda_t)^2|\lambda_t] + E[(n_t\lambda_t - n_t\lambda)^2|\lambda_t] \\
&= n_t\lambda_t(1-\lambda_t) + n_t^2(\lambda_t - \lambda)^2,
\end{aligned}$$

and then take expectations of this with respect to the unconditional distribution of λ_t:

$$\begin{aligned}
E(y_t - n_t\lambda)^2 &= n_t[E(\lambda_t) - E(\lambda_t^2)] + n_t^2 E(\lambda_t - \lambda)^2 \\
&= n_t(\lambda - \gamma_0 - \lambda^2) + n_t^2\gamma_0.
\end{aligned}$$

The unconditional covariance of y_t with y_{t-v} is likewise found from

$$\begin{aligned}
& E[(y_t - n_t\lambda)(y_{t-v} - n_{t-v}\lambda)|\lambda_t, \lambda_{t-v}] \\
= \; & E[(y_t - n_t\lambda_t + n_t\lambda_t - n_t\lambda)(y_{t-v} - n_{t-v}\lambda_{t-v} + n_{t-v}\lambda_{t-v} - n_{t-v}\lambda|\lambda_t, \lambda_{t-v}] \\
= \; & E[(y_t - n_t\lambda_t)(y_{t-v} - n_{t-v}\lambda_{t-v})|\lambda_t, \lambda_{t-v}] + E[(n_t\lambda_t - n_t\lambda)(n_{t-v}\lambda_{t-v} - n_{t-v}\lambda)|\lambda_t, \lambda_{t-v}] \\
= \; & 0 + n_t n_{t-v}(\lambda_t - \lambda)(\lambda_{t-v} - \lambda)
\end{aligned}$$

with unconditional expectation

$$E(y_t - n_t\lambda)(y_{t-v} - n_{t-v}\lambda) = n_t n_{t-v}\gamma_v.$$

The proposal is to estimate the unconditional probability of employment λ by maximizing the quasi likelihood $\ell(\lambda) = \sum_{t=1}^T \ell_t(\lambda)$ for $\ell_t(\lambda) = y_t \log \lambda + (n_t - y_t)\log(1-\lambda)$, from which the QMLE is calculated to be

$$\hat{\lambda} = n^{-1} \textstyle\sum_{t=1}^T y_t. \tag{A4}$$

We see immediately that $E(\hat{\lambda}) = n^{-1}\sum_{t=1}^T n_t\lambda = \lambda$, so the QMLE is an unbiased estimate of λ. Notice also

$$\hat{\lambda} - \lambda = n^{-1} \textstyle\sum_{t=1}^T (y_t - n_t\lambda)$$

so the variance of $\hat{\lambda}$ is

$$\begin{aligned}
E(\hat{\lambda} - \lambda)^2 \; = \; & n^{-2}\Big\{\textstyle\sum_{t=1}^T E(y_t - n_t\lambda)^2 + 2\sum_{t=2}^T E(y_t - n_t\lambda)(y_{t-1} - n_{t-1}\lambda) + \\
& 2\textstyle\sum_{t=3}^T E(y_t - n_t\lambda)(y_{t-2} - n_{t-2}\lambda) + \cdots + 2E(y_T - n_T\lambda)(y_1 - n_1\lambda)\Big\}
\end{aligned}$$

where we saw above that

$$E(y_t - n_t\lambda)(y_{t-v} - n_{t-v}\lambda) = \begin{cases} n_t(\lambda - \lambda^2 - \gamma_0) + n_t^2\gamma_0 & \text{for } v = 0 \\ n_t n_{t-v}\gamma_v & \text{for } v = 1, 2, ... \end{cases}.$$

We require that the number of individuals sampled each month n_t does not vary too much across months. For example, suppose that $n^{-1}\sum_{t=1}^T n_t \to 1$ and that for fixed v, $n^{-1}\sum_{t=v+1}^T n_t n_{t-v} \to$

η_v.[40] If the autocovariances of λ_t are absolutely summable ($\sum_{v=0}^{\infty} |\gamma_j| < \infty$), then

$$nE(\hat{\lambda} - \lambda)^2 \to V = \tilde{\gamma}_0 + 2\sum_{v=1}^{\infty} \tilde{\gamma}_v$$

for

$$\tilde{\gamma}_0 = (\lambda - \lambda^2 - \gamma_0) + \eta_0 \gamma_0$$

$$\tilde{\gamma}_v = \eta_v \gamma_v \quad \text{for } v = 1, 2, ...$$

Thus the variance of $\hat{\lambda}$ goes to zero as $n \to \infty$, confirming that $\hat{\lambda}$ is a consistent estimate of λ with

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, V).$$

We can also confirm that (A3) gives a consistent estimate of V. Here

$$h_t(\lambda) = \frac{y_t}{\lambda} - \frac{n_t - y_t}{(1-\lambda)} = \frac{y_t - n_t \lambda}{\lambda(1-\lambda)}$$

$$\frac{\partial h_t(\lambda)}{\partial \lambda} = -\frac{y_t}{\lambda^2} - \frac{(n_t - y_t)}{(1-\lambda)^2}$$

$$\hat{D} = n^{-1} \left\{ \frac{\sum_{t=1}^{T} y_t}{\lambda^2} + \frac{\sum_{t=1}^{T} (n_t - y_t)}{(1-\lambda)^2} \right\} \xrightarrow{p} \frac{\lambda}{\lambda^2} + \frac{(1-\lambda)}{(1-\lambda)^2} = \frac{1}{\lambda(1-\lambda)}$$

$$\hat{\Gamma}_v = n^{-1} \sum_{t=1}^{T} \left[\frac{y_t - n_t \lambda}{\lambda(1-\lambda)} \right] \left[\frac{y_{t-v} - n_{t-v} \lambda}{\lambda(1-\lambda)} \right] \xrightarrow{p} \frac{\tilde{\gamma}_v}{\lambda^2(1-\lambda)^2}$$

$$\hat{S} \xrightarrow{p} \frac{\tilde{\gamma}_0 + 2\sum_{v=1}^{\infty} \tilde{\gamma}_v}{\lambda^2(1-\lambda)^2}$$

$$\hat{V} = (\hat{D} \hat{S}^{-1} \hat{D})^{-1} \xrightarrow{p} \tilde{\gamma}_0 + 2\sum_{v=1}^{\infty} \tilde{\gamma}_v = V$$

as desired.

Appendix D. Weibull distribution.

Replacing mixture of exponentials (17)-(18) *with a Weibull distribution.* Suppose that a fraction of the population ω is newly unemployed each week with each person characterized by the same Weibull(α, λ) hazard rate of exiting unemployment. In steady state the fraction of the population

[40] If $n_t = \bar{n}$ for all t, these conditions are trivially satisfied with $n^{-1} \sum_{t=1}^{T} n_t = (T\bar{n})^{-1} \sum_{t=1}^{T} \bar{n} = 1$ and $n^{-1} \sum_{t=v+1}^{T} n_t n_{t-v} = (T\bar{n})^{-1} \sum_{t=v+1}^{T} \bar{n}^2 = \bar{n}(T - v)/T$.

unemployed for exactly τ weeks would then be given by[41]

$$\pi_U^{\dagger}(\tau) = \omega \exp(-\lambda \tau^{\alpha}). \tag{A5}$$

We accordingly maximized expression (20) with respect to $\theta_A, \pi_E, \pi_N, \pi_M, \omega, \alpha, \lambda$. This produced estimates $\hat{\alpha} = 0.404$ and $\hat{\lambda} = 0.861$, with other parameters similar to those for the baseline case in the first column of Table 1. The low value for α implies that unemployment-exit probabilities drop off significantly as the duration of unemployment τ increases. The Weibull captures the same basic features of the data as our baseline model. Nevertheless, the mixture of exponentials has a much better fit to the data, achieving a value for (20) that is substantially higher than that for the Weibull specification.[42]

Characterizing inconsistency between reported unemployment durations and observed hazards using the Weibull distribution. If we replace the unemployment-continuation probabilities predicted from the mixture of exponentials (17)-(18) by those predicted by the Weibull representation (A5) of the cross-section of unemployment durations, we would expect an average monthly unemployment-continuation probability of

$$\frac{\int_0^\infty \exp[-\lambda(\tau + 4.33)^{\alpha}]d\tau}{\int_0^\infty \exp[-\lambda\tau^{\alpha}]d\tau} = 0.67. \tag{A6}$$

This is far larger than the observed average monthly continuation probability across all unemployed of

$$T^{-1} \sum_{t=1}^{T} \left[\frac{\sum_{j \in J} y_{UU,t}^{[j]}}{\sum_{j \in J} \left(y_{UE,t}^{[j]} + y_{UN,t}^{[j]} + y_{UU,t}^{[j]} \right)} \right] = 0.55. \tag{A7}$$

Alternatively, our characterization of the cross-section implies an average monthly continuation probability across all individuals of $w_1 p_1^{4.33} + w_2 p_2^{4.33} = 0.70$. Whether one uses our parametric model, the Weibull, or any other, any model of the cross-section is calculating a parametric estimate of the magnitude

$$\frac{\sum_{\tau=5}^{99} y_{U,t}^{[j]}(\tau)}{\sum_{\tau=1}^{99} y_{U,t-1}^{[j]}(\tau)}$$

---

[41] The exponential distribution is a special case with $\alpha = 1$, $\exp(-\lambda) = p$ and $\omega = w(1-p)$.

[42] One might be tempted to think of twice this number (2,164) as a likelihood ratio statistic for testing whether the one extra parameter used by the mixture of exponentials is helpful. This number does not in fact have a $\chi^2(1)$ interpretation due to the strong serial correlation of $\ell_t(\lambda_X)$ and because the models are non-nested. Nevertheless, the huge magnitude of the difference in the quasi log likelihoods suggests that the baseline model is to be preferred.

which as we noted in Panel A of Figure 1 averages 0.72 over the sample. Any model that accurately describes the cross-section of durations– and ours does so quite well– is going to predict an unemployment-continuation probability similar to the stock-based measure plotted as the solid line in Panel A of Figure 1.

Appendix E. Details of modeling rotation bias.

For an individual who reported status $X^{[j]}$ in rotation j in month t, consider the counterfactual answer that individual would have given if interviewed using the interview technology that was used for rotation 1:

$$r^{[j]}_{X^{[j]},X^{[1]},t} = \text{ Prob(would have answered } X^{[1]} \text{ using technology 1 given answered } X^{[j]} \text{ using technology } j).$$

Collect these counterfactual probabilities in a matrix

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

Notice that each column of $R_t^{[j]}$ sums to unity. For example, for the first column, given that an individual reported status E when interviewed in rotation j, that person would have to have given one of the answers E, N, M, U if interviewed using the technology of rotation 1. We can construct a matrix $R_t^{[j]}$ that exactly satisfies the condition (6). For example, the first row states

$$r^{[j]}_{EE,t}\pi^{[j]}_{E,t} + r^{[j]}_{NE,t}\pi^{[j]}_{N,t} + r^{[j]}_{ME,t}\pi^{[j]}_{M,t} + r^{[j]}_{UE,t}\pi^{[j]}_{U,t} = \pi^{[1]}_{E,t}.$$

This equation states that the fraction who reported E in rotation 1 can be viewed as the fraction who reported $X^{[j]}$ in rotation j times the probability someone reporting $X^{[j]}$ would have reported E using technology 1, added across the four possible $X^{[j]}$. From the analysis in Section 3.1, for $j > 1$ we expect $r^{[j]}_{NU,t} > 0$; some of the individuals who report labor status N in rotation j would have reported status U if they had been interviewed for the first time. We also expect $r^{[j]}_{EM,t} > 0$ and $r^{[j]}_{NM,t} > 0$; some of the individuals who were reported as status E or N in rotation j would have been missing using the interview technology of rotation 1.

One can parameterize a matrix $R_t^{[j]}$ that exactly satisfies (6) in many ways. To construct monthly estimates adjusted for rotation bias we take the view that rotation bias evolves slowly over time, leading us to replace $R_t^{[j]}$ with an estimate $\bar{R}_t^{[j]}$ where $\bar{R}_t^{[j]}$ does not differ too much from $\bar{R}_{t-1}^{[j]}$. In this case, while (6) will always hold by construction when using $R_t^{[j]}$, if we instead use $\bar{R}_t^{[j]}$, then $\bar{R}_t^{[j]}\pi_t^{[j]}$ will be close to but not exactly equal to $\pi_t^{[1]}$. In anticipation of this objective, we parameterized the matrix $R_t^{[j]}$ in a way that focuses on what we believe to be the most important features of rotation bias. In Figure 2 we saw that the decline in U across rotations is balanced by a corresponding increase across rotations in N and that increases in M in rotations 1 and 5 match drops in E and N. We therefore propose to summarize the interview technology for rotation j in month t using three parameters $\theta_t^{[j]} = (\theta_{EM,t}^{[j]}, \theta_{NM,t}^{[j]}, \theta_{NU}^{[j]})'$:[43]

$$R_t^{[j]} = \begin{bmatrix} 1-\theta_{EM,t}^{[j]} & 0 & 0 & 0 \\ 0 & 1-\theta_{NM,t}^{[j]}-\theta_{NU,t}^{[j]} & 0 & 0 \\ \theta_{EM,t}^{[j]} & \theta_{NM,t}^{[j]} & 1 & 0 \\ 0 & \theta_{NU,t}^{[j]} & 0 & 1 \end{bmatrix}. \tag{A8}$$

The value of $\theta_t^{[j]}$ that causes (6) to hold exactly for every j is given by[44]

$$1-\theta_{EM,t}^{[j]} = \pi_{E,t}^{[1]}/\pi_{E,t}^{[j]} \tag{A9}$$

$$\theta_{NU,t}^{[j]} = (\pi_{U,t}^{[1]} - \pi_{U,t}^{[j]})/\pi_{N,t}^{[j]} \tag{A10}$$

[43] We take the (3,3) and (4,4) elements of $R_t^{[j]}$ to be unity because a higher fraction of the population is M or U in rotation 1 than in other rotations. For example, the third equation in (6) states that the fraction missing in rotation 1 is the fraction missing in rotation j plus some portions $\theta_{EM,t}^{[j]}$ and $\theta_{NM,t}^{[j]}$ of the fractions that are E and N in rotation j: $\pi_{M,t}^{[1]} = \pi_{M,t}^{[j]} + \theta_{EM,t}^{[j]}\pi_{E,t}^{[j]} + \theta_{NM,t}^{[j]}\pi_{N,t}^{[j]}$. Note that the normalization of the third and fourth columns of $R_t^{[j]}$ still allows equation (6) to fit exactly the observed values of every element of $\pi_t^{[j]}$ for every j and t.

[44] These equations come from solving rows 1 ,2 and 4 of (6), One can show that equations (A9)-(A11) imply that row 3 of (6) also holds. Add rows 1, 2, and 4 of (6) together to deduce

$$\pi_{E,t}^{[j]} + \pi_{U,t}^{[j]} + \pi_{N,t}^{[j]} - \theta_{EM,t}^{[j]}\pi_{E,t}^{[j]} - \theta_{NM,t}^{[j]}\pi_{N,t}^{[j]} = \pi_{E,t}^{[1]} + \pi_{U,t}^{[1]} + \pi_{N,t}^{[1]}.$$

Subtracting both sides from 1 gives

$$\pi_{M,t}^{[j]} + \theta_{EM,t}^{[j]}\pi_{E,t}^{[j]} + \theta_{NM,t}^{[j]}\pi_{N,t}^{[j]} = \pi_{M,t}^{[1]}$$

as required by the third row of (6). In general, since each column of $R_t^{[j]}$ sums to unity, if elements of $\pi_t^{[j]}$ sum to unity, then the elements of $R_t^{[j]}\pi_t^{[j]}$ also sum to unity: $1'R_t^{[j]}\pi_t = 1'\pi_t = 1$ for 1 a vector of four ones.

$$1 - \theta^{[j]}_{NM,t} - \theta^{[j]}_{NU,t} = \pi^{[1]}_{N,t} / \pi^{[j]}_{N,t}. \tag{A11}$$

Let $\Pi^{[j]}$ be the observed full-sample average transition probabilities into rotation j and $\bar{R}^{[j]}$ be the value obtained by plugging the parameter values in Table 2 into expression (A8). We then chose values for the $(n \times n)$ matrix Π^* by minimizing the sum of squared elements of

$$\Pi^{[j]} - (\bar{R}^{[j]})^{-1} \Pi^* \bar{R}^{[j-1]} \quad \text{for } j \in J \tag{A12}$$

$$\pi^{[1]} - \pi^* \tag{A13}$$

$$\pi^{[5]} - (\bar{R}^{[5]})^{-1} \pi^* \tag{A14}$$

subject to the constraints that all elements of Π^* lie between 0 and 1, each column of Π^* sums to 1, and that π^* is the vector of ergodic probabilities implied by Π^*.[45]

This framework predicts that the fraction of individuals reporting status E, N, M, or U when interviewed using technology j would be given by

$$\hat{\pi}^{[j]} = (\bar{R}^{[j]})^{-1} \pi^*. \tag{A15}$$

These predicted values $\hat{\pi}^{[j]}$ were compared with the actual values $\pi^{[j]}$ in Figure 3. Our approach also implies a predicted value for the observed fraction of individuals with measured transitions from $X^{[j-1]}$ to $X^{[j]}$:

$$\hat{\Pi}^{[j]} = (\bar{R}^{[j]})^{-1} \Pi^* \bar{R}^{[j-1]}. \tag{A16}$$

The predicted $\hat{\Pi}^{[j]}$ are compared with the observed $\Pi^{[j]}$ in Figure 4.

To construct month-by-month estimates, our first step is to construct weighted moving averages of the counts of individuals in each labor-force status in each rotation as in (9). We then calculated

[45] That is, we minimized the sum of squares of the $96 = 16 \times 6$ elements in (A12) plus the sum of squares of the 8 elements in (A13) and (A14). The vector π^* is also a function of Π^* using expression [22.2.26] in Hamilton (1994):

$$B = \begin{bmatrix} I_4 - \Pi^* \\ 1' \end{bmatrix}$$

$$\pi^* = (B'B)^{-1} B' e_5$$

where $1'$ denotes a (1×4) vector of ones and e_5 denotes column 5 of I_5.

the corresponding smoothed fractions as

$$\overline{\pi}_{X,t}^{[j]} = \overline{y}_{X,t}^{[j]} / \left(\overline{y}_{E,t}^{[j]} + \overline{y}_{N,t}^{[j]} + \overline{y}_{M,t}^{[j]} + \overline{y}_{U,t}^{[j]} \right).$$

From these we calculated time-varying rotation-bias parameters as

$$\theta_{EM,t}^{[j]} = \max \left\{ 1 - \left(\overline{\pi}_{E,t}^{[1]} / \overline{\pi}_{E,t}^{[j]} \right), 0 \right\}$$

$$\theta_{NU,t}^{[j]} = \max \left\{ \left(\overline{\pi}_{U,t}^{[1]} - \overline{\pi}_{U,t}^{[j]} \right) / \overline{\pi}_{N,t}^{[j]}, 0 \right\}$$

$$\theta_{NM,t}^{[j]} = \max \left\{ 1 - \theta_{NU,t}^{[j]} - \left(\overline{\pi}_{N,t}^{[1]} / \overline{\pi}_{N,t}^{[j]} \right), 0 \right\},$$

Plugging the values for $\overline{\theta}_{EM,t}^{[j]}$, $\overline{\theta}_{NU,t}^{[j]}$, and $\overline{\theta}_{NM,t}^{[j]}$ into (A8) gives a value of $\bar{R}_t^{[j]}$ for each j and t. Our procedure was to proceed iteratively through the data, choosing Π_t^* for each t to minimize the errors in the following equations:

$$\Pi_t^{[j]} - (\bar{R}_t^{[j]})^{-1} \Pi_t^* \bar{R}_{t-1}^{[j-1]} \qquad \text{for } j \in J = \{2, 3, 4\} \cup \{6, 7, 8\} \tag{A17}$$

$$\pi_t^{[1]} - \Pi_t^* \pi_{t-1}^* \tag{A18}$$

$$\pi_t^{[5]} - (\bar{R}_t^{[5]})^{-1} \Pi_t^* \pi_{t-1}^*. \tag{A19}$$

We set the initial value of π_t^* for observation $t = 1$ as $\pi_1^* = (\pi_1^{[1]} + \pi_1^{[5]})/2$. For each $t = 2, 3, ...$ we choose the 16 elements of Π_t^* so as to minimize the sum of squares of the 104 terms in (A17)-(A19) subject to the constraints that each element of Π_t^* is between 0 and 1 and each column of Π_t^* sums to 1. Given Π_t^* we then calculated

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

and proceeded to the next observation $t + 1$.

Table A-1. Summary of notation

$\pi_X^{[j]}$ = fraction of working-age population reporting status $X \in \{E, N, M, U\}$ in rotation $j \in \{1, ..., 8\}$

$\pi_{X_1,X_2}^{[j]}$ = fraction of population reporting X_1 in rotation $j-1$ and X_2 in rotation $j \in \{2, ..., 8\}$

$y_{X,t}^{[j]}$ = weighted number of people reporting status $X \in \{E, N, M, U\}$ in rotation j in month t

J = the set consisting of rotations $\{2, 3, 4\} \cup \{6, 7, 8\}$

$y_{X_1,X_2,t}^{[j]}$ = weighted number reporting X_1 in rotation $j-1$ in month $t-1$ and X_2 in rotation $j \in J$ in month t

$y_{U,t}^{[j]}(\tau)$ = weighted number in rotation j in month t reporting U with duration τ

$y_{X,U,t}^{[j]}(\tau)$ = weighted number reporting $X \in \{E, N, M\}$ in rotation $j-1$ in month $t-1$ and U with duration τ in month t for $j \in J$

$y_{U,X,t}^{[j]}(\tau)$ = weighted number reporting U with duration τ in rotation $j-1$ in month $t-1$ and reporting $X \in \{E, N, M, U\}$ in month t for $j \in \{2, ..., 8\}$

p_i = weekly unemployment-continuation probability consistent with reported unemployment duration for type $i \in \{1, 2\}$

w_i = fraction of unemployed who are type $i \in \{1, 2\}$

$\dot{\pi}_U(\tau)$ = predicted fraction of unemployed who report duration τ

θ_A = vector of parameters characterizing matrix A of rounding errors in reporting durations

$\pi_U^{\dagger}(\tau)$ = imputed fraction of unemployed with perceived duration τ in absence of rounding errors

$\dot{\pi}_{X,U}(\tau)$ = of the people who report status $X \in \{E, N, M\}$ in $t-1$ and U in t, the predicted fraction who report duration τ

$q_{i,XU}$ = of the people who report status $X \in \{E, N, M, U\}$ in $t-1$ and U in t, the fraction who report duration i for $i \in \{1, ..., 4\}$ or report duration greater than 4 weeks with perceived type 1 or type 2 duration for $i \in \{5, 6\}$

$\eta_i(\tau)$ = probability an individual is type $i \in \{1, 2\}$ given they report duration $\tau \in \{1, ..., 99\}$

$\gamma_{i,UX}$ = probability an individual of type i who is unemployed in month $t-1$ will report status $X \in \{E, N, M, U\}$ in month t

$\theta_{EM}^{[j]}, \theta_{NU}^{[j]}, \theta_{NM}^{[j]}$ = parameters characterizing rotation bias for rotation $j \in \{2, ..., 8\}$

π_X^* = fraction of population with reported status $X \in \{E, N, M, U\}$ after correcting for rotation bias

$\pi^* = (4 \times 1)$ vector containing $(\pi_E^*, \pi_N^*, \pi_M^*, \pi_U^*)'$

π_{X_1,X_2}^* = probability of reporting status X_2 in month t conditional on reporting X_1 in month $t-1$ after correcting for rotation bias

$\Pi^* = (4 \times 4)$ matrix collecting the values of π_{X_1,X_2}^*

$\tilde{\pi}_X$ = fraction of population inferred to have true status $X \in \{E, N, U\}$ after correcting for rotation bias, nonrandom missing observations, and misclassified N

Table A-2. Cohorts affected by start of sample in July 2001.

MIS in 2001:7	7	6	5	x	x	x	x	x	x	x	x	4	4	3	2	1
Apr-00																
May-00	E1															
Jun-00	E2	E1														
Jul-00	E3	E2	E1													
Aug-00	E4	E3	E2	E1												
Sep-00	x	E4	E3	E2	E1											
Oct-00	x	x	E4	E3	E2	E1										
Nov-00	x	x	x	E4	E3	E2	E1									
Dec-00	x	x	x	x	E4	E3	E2	E1								
Jan-01	x	x	x	x	x	E4	E3	E2	E1							
Feb-01	x	x	x	x	x	x	E4	E3	E2	E1						
Mar-01	x	x	x	x	x	x	x	E4	E3	E2	E1					
Apr-01	x	x	x	x	x	x	x	x	E4	E3	E2	E1	E1			
May-01	E5	x	x	x	x	x	x	x	x	E4	E3	E2	E2	E1		
Jun-01	E6	E5	x	x	x	x	x	x	x	x	E4	E3	E3	E2	E1	
Jul-01	M7	M6	M5	x	x	x	x	x	x	x	x	E4	M4	M3	M2	E1
Aug-01	M8	M7	M6	M5	x	x	x	x	x	x	x	x	x	M4	M3	M2
Sep-01		M8	M7	M6	M5	x	x	x	x	x	x	x	x	x	M4	M3
Oct-01			M8	M7	M6	M5	x	x	x	x	x	x	x	x	x	M4
Nov-01				M8	M7	M6	M5	x	x	x	x	x	x	x	x	x
Dec-01					M8	M7	M6	M5	x	x	x	x	x	x	x	x
Jan-02						M8	M7	M6	M5	x	x	x	x	x	x	x
Feb-02							M8	M7	M6	M5	x	x	x	x	x	x
Mar-02								M8	M7	M6	M5	x	x	x	x	x
Apr-02									M8	M7	M6	M5	M5	x	x	x
May-02										M8	M7	M6	M6	M5	x	x
Jun-02											M8	M7	M7	M6	M5	x
Jul-02												M8	M8	M7	M6	M5
Aug-02														M8	M7	M6
Sep-02															M8	M7
Oct-02																M8
Nov-02																

Table A-3. Cohorts affected by end of sample (using April 2018 to illustrate sample end).

MIS in 18:4		7	6	5	x	x	x	x	x	x	x	x	4	3	2	1
Jan-17	M1															
Feb-17	M2	M1														
Mar-17	M3	M2	M1													
Apr-17	M4	M3	M2	M1												
May-17	x	M4	M3	M2	M1											
Jun-17	x	x	M4	M3	M2	M1										
Jul-17	x	x	x	M4	M3	M2	M1									
Aug-17	x	x	x	x	M4	M3	M2	M1								
Sep-17	x	x	x	x	x	M4	M3	M2	M1							
Oct-17	x	x	x	x	x	x	M4	M3	M2	M1						
Nov-17	x	x	x	x	x	x	x	M4	M3	M2	M1					
Dec-17	x	x	x	x	x	x	x	x	M4	M3	M2	M1				
Jan-18	M5	x	x	x	x	x	x	x	x	M4	M3	M2	M1			
Feb-18	M6	M5	x	x	x	x	x	x	x	x	M4	M3	M2	M1		
Mar-18	M7	M6	M5	x	x	x	x	x	x	x	x	M4	M3	M2	M1	
Apr-18	E8	M7	M6	M5	x	x	x	x	x	x	x	x	M4	M3	M2	M1
May-18		E8	M7	M6	M5	x	x	x	x	x	x	x	x	M4	M3	M2
Jun-18			E8	M7	M6	M5	x	x	x	x	x	x	x	x	M4	M3
Jul-18				E8	M7	M6	M5	x	x	x	x	x	x	x	x	M4
Aug-18					E8	M7	M6	M5	x	x	x	x	x	x	x	x
Sep-18						E8	M7	M6	M5	x	x	x	x	x	x	x
Oct-18							E8	M7	M6	M5	x	x	x	x	x	x
Nov-18								E8	M7	M6	M5	x	x	x	x	x
Dec-18									E8	M7	M6	M5	x	x	x	x
Jan-19										E8	M7	M6	M5	x	x	x
Feb-19											E8	M7	M6	M5	x	x
Mar-19												E8	M7	M6	M5	x
Apr-19													E8	M7	M6	M5
May-19														E8	M7	M6
Jun-19															E8	M7
Jul-19																E8
Aug-19																

Table A-4. Cohorts affected by sample redesign in August 2004.

May-03													
Jun-03													
Jul-03													
Aug-03	M1												
Sep-03	M2	M1											
Oct-03	M3	M2	M1										
Nov-03	M4	M3	M2	M1									
Dec-03	x	M4	M3	M2	M1								
Jan-04	x	x	M4	M3	M2	M1							
Feb-04	x	x	x	M4	M3	M2	M1						
Mar-04	x	x	x	x	M4	M3	M2	M1					
Apr-04	x	x	x	x	x	M4	M3	M2	M1				
May-04	x	x	x	x	x	x	M4	M3	M2	M1			
Jun-04	x	x	x	x	x	x	x	M4	M3	M2	M1		
Jul-04	x	x	x	x	x	x	x	x	M4	M3	M2	M1	
Aug-04	E5	x	x	x	x	x	x	x	x	M4	M3	M2	E1
Sep-04	E6	E5	x	x	x	x	x	x	x	x	M4	M3	E2
Oct-04	E7	E6	E5	x	x	x	x	x	x	x	x	M4	E3
Nov-04	E8	E7	E6	E5	x	x	x	x	x	x	x	x	E4
Dec-04		E8	E7	E6	E5	x	x	x	x	x	x	x	x
Jan-05			E8	E7	E6	E5	x	x	x	x	x	x	x
Feb-05				E8	E7	E6	E5	x	x	x	x	x	x
Mar-05					E8	E7	E6	E5	x	x	x	x	x
Apr-05						E8	E7	E6	E5	x	x	x	x
May-05							E8	E7	E6	E5	x	x	x
Jun-05								E8	E7	E6	E5	x	x
Jul-05									E8	E7	E6	E5	x
Aug-05										E8	E7	E6	E5
Sep-05											E8	E7	E6
Oct-05												E8	E7
Nov-05													E8

Table A-4 (continued).

May-03	E1														
Jun-03	E2	E1													
Jul-03	E3	E2	E1												
Aug-03	E4	E3	E2	E1											
Sep-03	x	E4	E3	E2	E1										
Oct-03	x	x	E4	E3	E2	E1									
Nov-03	x	x	x	E4	E3	E2	E1								
Dec-03	x	x	x	x	E4	E3	E2	E1							
Jan-04	x	x	x	x	x	E4	E3	E2	E1						
Feb-04	x	x	x	x	x	x	E4	E3	E2	E1					
Mar-04	x	x	x	x	x	x	x	E4	E3	E2	E1				
Apr-04	x	x	x	x	x	x	x	x	E4	E3	E2	E1			
May-04	E5	x	x	x	x	x	x	x	x	E4	E3	E2	E1		
Jun-04	E6	E5	x	x	x	x	x	x	x	x	E4	E3	E2	E1	
Jul-04	E7	E6	E5	x	x	x	x	x	x	x	x	E4	E3	E2	E1
Aug-04	E8	E7	E6	M5	x	x	x	x	x	x	x	x	E4	E3	E2
Sep-04		E8	E7	M6	M5	x	x	x	x	x	x	x	x	E4	E3
Oct-04			E8	M7	M6	M5	x	x	x	x	x	x	x	x	E4
Nov-04				M8	M7	M6	M5	x	x	x	x	x	x	x	x
Dec-04					M8	M7	M6	M5	x	x	x	x	x	x	x
Jan-05						M8	M7	M6	M5	x	x	x	x	x	x
Feb-05							M8	M7	M6	M5	x	x	x	x	x
Mar-05								M8	M7	M6	M5	x	x	x	x
Apr-05									M8	M7	M6	M5	x	x	x
May-05										M8	M7	M6	M5	x	x
Jun-05											M8	M7	M6	M5	x
Jul-05												M8	M7	M6	M5
Aug-05													M8	M7	M6
Sep-05														M8	M7
Oct-05															M8
Nov-05															

Table A-5. Estimated average fractions of individuals π_X^{**} who would have reported labor status *E, N, M*, or *U* and transition probabilities π_{X_1,X_2}^{**} if all individuals were being interviewed using the average interview technology.

$$\begin{bmatrix}\pi_E^{**}\\ \pi_N^{**}\\ \pi_M^{**}\\ \pi_U^{**}\end{bmatrix}=\begin{bmatrix}0.4257\\ 0.2499\\ 0.2966\\ 0.0278\end{bmatrix} \qquad \begin{bmatrix}\pi_{EE}^{**} & \pi_{NE}^{**} & \pi_{ME}^{**} & \pi_{UE}^{**}\\ \pi_{EN}^{**} & \pi_{NN}^{**} & \pi_{MN}^{**} & \pi_{UN}^{**}\\ \pi_{EM}^{**} & \pi_{NM}^{**} & \pi_{MM}^{**} & \pi_{UM}^{**}\\ \pi_{EU}^{**} & \pi_{NU}^{**} & \pi_{MU}^{**} & \pi_{UU}^{**}\end{bmatrix}=\begin{bmatrix}0.8967 & 0.0407 & 0.0945 & 0.2072\\ 0.0274 & 0.8705 & 0.0501 & 0.2096\\ 0.0634 & 0.0647 & 0.8470 & 0.0791\\ 0.0124 & 0.0241 & 0.0084 & 0.5040\end{bmatrix}$$

Figure A-1. Probability of *MM* transitions before and after adjustments (July 2001 to February 2020)

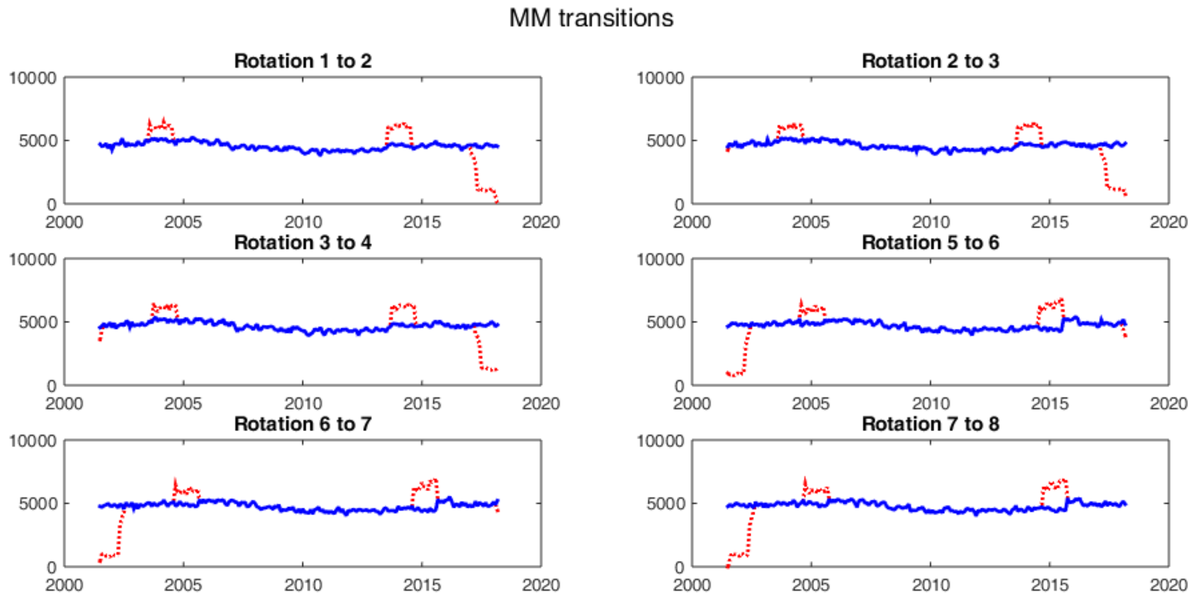
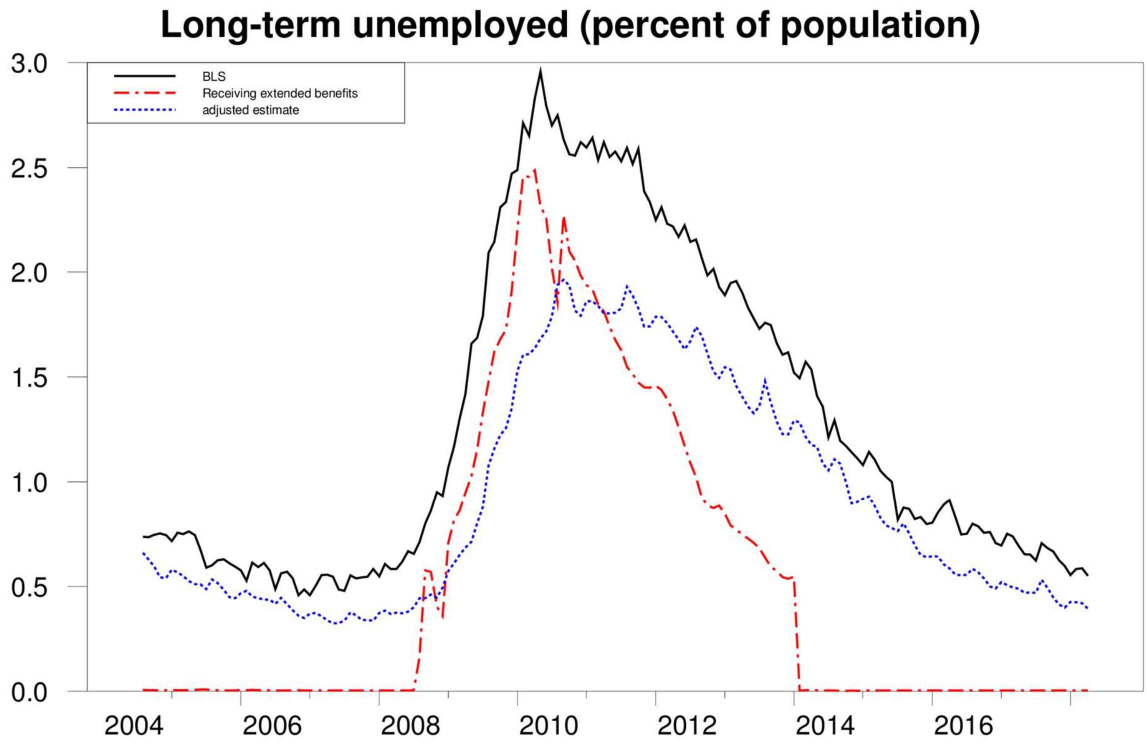


Figure A-2. Alternative measures of number of long-term unemployed as a percent of the civilian noninstitutional population.



Notes to Figure A-2. Long-term unemployed as percent of the civilian noninstitutional population, Aug 2004 to April 2018. Solid black: BLS estimate of number of unemployed with durations 27 weeks and over; Dashed red: number of individuals collecting Emergency Unemployment Compensation 2008; dotted blue: adjusted estimate based on equation (34).