Appendix to
"Underemployment Following the Great Recession and the COVID-19 Recession"
Daniela Dean Avila and Kurt G. Lunsford

Appendix A: Testing for a Structural Break

In this appendix, we test for a structural break in the relationship between the underemployment rate and unemployment rate. In particular, we test for a break in the intercept of a linear conditional expectation model that would be consistent with the level shift plotted in Figure 2 of the *Commentary*. We find evidence of a structural beak in January 2008. This timing aligns with the beginning of the 2008–2009 Great Recession.

Let y_t denote the underemployment rate in month t and u_t denote the unemployment rate in month t. We index the available data sample with $t = 1, \ldots, T$. We estimate the expectation of y_t conditional on u_t, using a linear model

$$y_t = \alpha_t + u_t\beta + e_t. \tag{A.1}$$

The parameter β is the slope coefficient, which we assume to be constant. The parameter α_t is the intercept which we allow to be potentially time-varying. Following Andrews (1993), we test the null hypothesis of α_t being constant, $H_0\colon \alpha_t = \alpha_0$, against the alternative hypothesis of a one-time change,

$$H_a\colon \begin{cases} \alpha_t = \alpha_1, & \text{for } t = 1, \ldots, T_1 \\ \alpha_t = \alpha_2, & \text{for } t = T_1 + 1, \ldots, T \end{cases}$$

for some T_1 with $1 < T_1 < T$. We will test this null hypothesis against the alternative for different values of T_1 by computing a Wald statistic for each choice of T_1.

First, we estimate the parameters α_t and β. If we did not permit time variation in α_t, then we could estimate α and β in (A.1) with a generalized method of moments (GMM) approach, using $E(e_t) = 0$ and $E(u_t e_t) = 0$ as our two moments to identify α and β. Using (A.1), these two moments can be written as $E(y_t - \alpha - u_t\beta) = 0$ and $E\big(u_t(y_t - \alpha - u_t\beta)\big) = 0$. The GMM estimator would then minimize $\bar{m}(\alpha,\beta)'W\bar{m}(\alpha,\beta)$, in which

$$\bar{m}(\alpha,\beta) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \begin{bmatrix} y_t - \alpha - u_t\beta \\ u_t(y_t - \alpha - u_t\beta) \end{bmatrix}$$

and W is a (2×2) positive definite weighting matrix. Instead, for a given choice of T_1, we allow $\alpha_t = \alpha_1$ for $t = 1, \ldots, T_1$ and $\alpha_t = \alpha_2$ for $t = T_1 + 1, \ldots, T$. Our GMM estimator minimizes $\bar{m}(\alpha_1, \alpha_2, \beta, T_1)'W\bar{m}(\alpha_1, \alpha_2, \beta, T_1)$, in which

$$\bar{m}(\alpha_1, \alpha_2, \beta, T_1) = \frac{1}{T}\sum_{t=1}^{T_1} \begin{bmatrix} y_t - \alpha_1 - u_t\beta \\ u_t(y_t - \alpha_1 - u_t\beta) \\ 0 \\ 0 \end{bmatrix} + \frac{1}{T}\sum_{t=T_1+1}^{T} \begin{bmatrix} 0 \\ 0 \\ y_t - \alpha_2 - u_t\beta \\ u_t(y_t - \alpha_2 - u_t\beta) \end{bmatrix} \tag{A.2}$$

and W is a (4×4) positive definite weighting matrix.

Next, we change the notation for the GMM problem and (A.2). We define

$$Y_1 = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{T_1} \end{bmatrix}, Y_2 = \begin{bmatrix} Y_{T_1+1} \\ \vdots \\ Y_T \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, U_1 = \begin{bmatrix} u_1 \\ \vdots \\ u_{T_1} \end{bmatrix}, U_2 = \begin{bmatrix} u_{T_1+1} \\ \vdots \\ u_T \end{bmatrix}$$

$$X = \begin{bmatrix} 1_{T_1\times 1} & 0_{T_1\times 1} & U_1 \\ 0_{T-T_1\times 1} & 1_{T-T_1\times 1} & U_2 \end{bmatrix}, Z = \begin{bmatrix} 1_{T_1\times 1} & U_1 & 0_{T_1\times 1} & 0_{T_1\times 1} \\ 0_{T-T_1\times 1} & 0_{T-T_1\times 1} & 1_{T-T_1\times 1} & U_2 \end{bmatrix}, \theta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}$$

in which $0_{m\times n}$ is a $(m \times n)$ matrix of 0s and $1_{m\times n}$ is a $(m \times n)$ matrix of 1s. Then, (A.2) can be written as

$$\overline{m}(\alpha_1, \alpha_2, \beta, T_1) = T^{-1}Z'(Y - X\theta). \tag{A.3}$$

In the GMM problem, we then use $W = (Z'Z)^{-1}$, and the solution to the minimization problem is

$$\hat{\theta} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta} \end{bmatrix} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y, \tag{A.4}$$

in which $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\beta}$ will take different values for different choices of T_1. We then compute

$$\hat{E} = \begin{bmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_T \end{bmatrix} = Y - X\hat{\theta}, \tag{A.5}$$

in which $\hat{e}_t$ for $t = 1, \ldots, T$ are the residuals.

Given these estimates for each choice of T_1, we can then compute a Wald statistic for each choice of T_1. The Wald statistic is

$$W = T(\hat{\alpha}_1 - \hat{\alpha}_2)\left(\frac{T}{T_1}\hat{V}_1 + \frac{T}{T - T_1}\hat{V}_2\right)^{-1}(\hat{\alpha}_1 - \hat{\alpha}_2), \tag{A.6}$$

in which $\hat{V}_1$ and $\hat{V}_2$ are variance estimates. To compute these variance estimates, we begin with some notation. Let

$$\widehat{m}_t = \begin{bmatrix} \hat{e}_t \\ u_t\hat{e}_t \end{bmatrix}$$

for $t = 1, \ldots, T$, $\overline{\widehat{m}}_1 = \sum_{t=1}^{T_1} \widehat{m}_t$, and $\overline{\widehat{m}}_2 = \sum_{t=T_1+1}^{T} \widehat{m}_t$. We denote the centered values of $\widehat{m}_t$ with $\hat{z}_{1,t} = \widehat{m}_t - \overline{\widehat{m}}_1$ for $t = 1, \ldots, T_1$ and $\hat{z}_{2,t} = \widehat{m}_t - \overline{\widehat{m}}_2$ for $t = T_1 + 1, \ldots, T$. Then, $\hat{V}_1$ and $\hat{V}_2$ are given by

$$\hat{V}_1 = \left(\widehat{M}_1'\widehat{\Omega}_1^{-1}\widehat{M}_1\right)^{-1} \text{and } \hat{V}_2 = \left(\widehat{M}_2'\widehat{\Omega}_2^{-1}\widehat{M}_2\right)^{-1}, \tag{A.7}$$

in which

$$\widehat{M}_1 = \frac{1}{T_1}\sum\nolimits_{t=1}^{T_1}\begin{bmatrix} -1 \\ -u_t \end{bmatrix}, \widehat{M}_2 = \frac{1}{T - T_1}\sum\nolimits_{t=T_1+1}^{T}\begin{bmatrix} -1 \\ -u_t \end{bmatrix}, \tag{A.8}$$

$$\widehat{\Omega}_1 = \frac{1}{T_1}\sum_{t=1}^{T_1} z_{1,t}z_{1,t}' + \frac{1}{T_1}\sum_{j=1}^{h_1-1}\sum_{t=1}^{T_1-j} k_{1,j}\left(z_{1,t}z_{1,t+j}' + z_{1,t+j}z_{1,t}'\right), \tag{A.9}$$

and

$$\widehat{\Omega}_2 = \frac{1}{T-T_1}\sum_{t=T_1+1}^{T} z_{2,t}z'_{2,t} + \frac{1}{T-T_1}\sum_{j=1}^{h_2-1}\sum_{t=T_1+1}^{T-j} k_{2,j}\left(z_{2,t}z'_{2,t+j} + z_{2,t+j}z'_{2,t}\right). \quad \text{(A.11)}$$

Following Lazarus et al. (2018), we set h_1 equal to $1.3T_1^{1/2}$, rounded to the nearest integer, and h_2 equal to $1.3(T-T_1)^{1/2}$, rounded to the nearest integer. Then, $k_{1,j}$ and $k_{2,j}$ are Bartlett kernels (Newey and West, 1987) given by $k_{1,j} = 1 - j/h_1$ and $k_{2,j} = 1 - j/h_2$.

The data sample runs from January 1994 to November 2021, giving $T = 335$. We do not permit a break in the first 48 or the last 48 observations of our sample. That is, we compute Wald statistics for each month from January 1998 to November 2017, and we do not formally test for a second break after the 2020 recession. Our sample of Wald statistics cuts off the first and last 14 percent of observations in the total sample. Then, the 1 percent critical value for rejecting the null hypothesis is roughly 12.16 (Andrews, 2003).

Figure 1 shows the Wald statistics from January 1998 to November 2017. The maximum Wald statistic occurs in January 2008 and has a value of 48.76. This value is well above the 1 percent critical value, providing evidence of a structural break in the intercept. Prior to the structural break, we find $\hat{\alpha}_1 = 0.20$. After the break, $\hat{\alpha}_2 = 0.97$. This difference implies an upward shift of 0.77. This value is slightly less than the 1 percent found in Valetta et al. (2020) and discussed in the *Commentary*. However, we note that our data sample includes the data after the 2020 COVID-19 recession, data which appear more consistent with the pre-2008 structure based on Figure 2 in the *Commentary* and may pull $\hat{\alpha}_2$ down slightly.

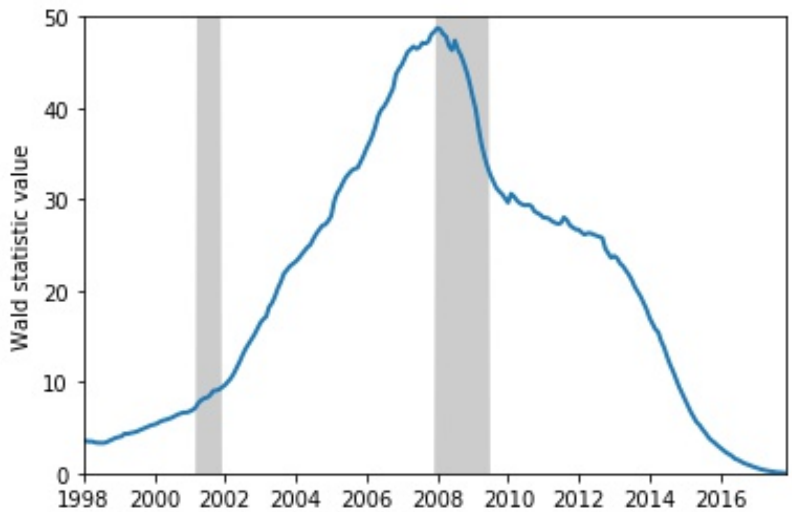

Figure 1: Wald statistics testing for intercept breaks from January 1998 to November 2017

Note: Gray bars indicate recessions.

Source: US Bureau of Labor Statistics, Employment Level: Part-Time for Economic Reasons [LNS12032194], Employment Level [CE16OV], Unemployment Level [UNEMPLOY], and Civilian Labor Force Level [CLF16OV], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org and authors' calculations.

Appendix B. Industry Decomposition of the Underemployment Rate

In this appendix, we provide a decomposition of the underemployment rate by industry. The underemployment rate is the level of underemployment divided by the level of employment times 100, $y_t = 100 \times Y_t/E_t$. Both the levels of employment and underemployment comprise J industries, indexed by $j = 1, \dots, J$. Then, we have

$$y_t = 100 \times \frac{Y_{1,t}}{E_{1,t}}\frac{E_{1,t}}{E_t} + 100 \times \frac{Y_{2,t}}{E_{2,t}}\frac{E_{2,t}}{E_t} + \cdots + 100 \times \frac{Y_{J,t}}{E_{J,t}}\frac{E_{J,t}}{E_t}. \quad \text{(B.1)}$$

We use the notation $y_{j,t} = 100 \times Y_{j,t}/E_{j,t}$ to denote the industry-specific underemployment rate and $w_{j,t} = E_{j,t}/E_t$ to denote the industry-specific employment share. Then, (B.1) is the same as

$$y_t = y_{1,t}w_{1,t} + y_{2,t}w_{2,t} + \cdots + y_{J,t}w_{J,t}. \quad \text{(B.2)}$$

Hence, the total underemployment rate is a weighted average of industry-specific underemployment rates, and the employment shares provide the weights. The change in the underemployment rate between period s and period t is given by

$$y_t - y_s = (y_{1,t}w_{1,t} - y_{1,s}w_{1,s}) + (y_{2,t}w_{2,t} - y_{2,s}w_{2,s}) + \cdots + (y_{J,t}w_{J,t} - y_{J,s}w_{J,s}). \quad \text{(B.3)}$$

For industry j, we have

$$y_{j,t}w_{j,t} - y_{j,s}w_{j,s} = y_{j,s}(w_{j,t} - w_{j,s}) + (y_{j,t} - y_{j,s})w_{j,s} + (y_{j,t} - y_{j,s})(w_{j,t} - w_{j,s}), \quad \text{(B.4)}$$

which is the decomposition that we will use to study how industries affect the underemployment rate. The first term on the right-hand side of (B.4) is the change in industry j's employment share times its initial underemployment rate. This term measures the change in the industry employment share while holding the underemployment rate fixed. The second term of the right-hand side of (B.4) is the change in industry j's underemployment rate times its initial employment share. This term measures the change in the industry unemployment rate while holding the employment share fixed. The third term is a cross term that we will ignore because it is generally very small in magnitude. Hence, combining (B.3) and (B.4), we have

$$y_t - y_s \approx \sum_{j=1}^{J} y_{j,s}(w_{j,t} - w_{j,s}) + \sum_{j=1}^{J} (y_{j,t} - y_{j,s})w_{j,s}, \quad \text{(B.5)}$$

which decomposes the change in the underemployment rate into industry employment share changes and industry underemployment rate changes.

References

Andrews, Donald W. K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Chang Point." *Econometrica* 61 (4): 821–56. https://doi.org/10.2307/2951764.

Andrews, Donald W. K. 2003. "Tests for Parameter Instability and Structural Change with Unknown Chang Point: A Corrigendum." *Econometrica* 71 (1): 395–97. https://doi.org/10.1111/1468-0262.00405.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson. 2018. "HAR Inference: Recommendations for Practice." *Journal Business & Economic Statistics* 36 (4): 541–59. https://doi.org/10.1080/07350015.2018.1506926.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708. https://doi.org/10.2307/1913610.

Valetta, Robert G., Leila Bengali, and Catherine van der List. 2020. "Cyclical and Market Determinants of Involuntary Part-Time Employment." *Journal of Labor Economics* 38 (1): 67-93. https://doi.org/10.1086/704496.