EC317 – Labour Economics

Problem Set 10 Solutions

1. Consider the following equation for log-hourly wages

$$\ln(w_{it}) = \alpha + \beta\, d_{it}^{\texttt{union}} + x_i'\theta + \varepsilon_{it}, \tag{1}$$

where w_{it} is worker i's hourly wage in wave $t \in \{1,5\}$, $d_{it}^{\texttt{union}}$ is a dummy indicating union membership in wave t, and x_i is a vector of time-invariant worker characteristics.

Obtain a cross-sectional estimate of the union wage differential β by OLS estimation of equation (1) for log-hourly wages in wave 1, controlling for the worker's age in wave 1 and its square, sex, and race. Weight your regression using the survey weights (variable `weight`) to account for the LFS stratified sampling design —type `help weight` and `help regress` in Stata to learn how to do this.

What is your estimate of the union wage differential?

Answer:

```
. regress lhw1 tu1 age1 age1sq sex white [pweight = weight], robust
(sum of wgt is 227,467,244.8949)

Linear regression                               Number of obs     =     33,635
                                                F(5, 33629)       =     713.18
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1473
                                                Root MSE          =     .51997

------------------------------------------------------------------------------
             |               Robust
        lhw1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         tu1 |   .1113967   .0074065    15.04   0.000     .0968798    .1259137
        age1 |   .0858845   .0020015    42.91   0.000     .0819614    .0898076
      age1sq |  -.0009723   .0000244   -39.85   0.000    -.0010201   -.0009245
         sex |  -.2183974   .0073544   -29.70   0.000    -.2328124   -.2039824
       white |   .0306437   .0195796     1.57   0.118     -.007733    .0690204
       _cons |   1.018546   .0438685    23.22   0.000     .9325616    1.104529
------------------------------------------------------------------------------
```

The point estimate $\widehat{\beta}^{\texttt{cross}} = 0.111$ implies cross-sectional union wage differential of 11.1%.

2. Now, consider the *first-differenced* equation

$$\Delta \log(w_i) = \gamma + \beta\, \Delta d_i^{\texttt{union}} + \Delta\varepsilon_i \tag{2}$$

where $\Delta(\cdot)_i \equiv (\cdot)_{i5} - (\cdot)_{i1}$, and the intercept γ allows for a linear time trend in the underlying *levels* equation.

Estimate equation (2) by OLS. Again, use the survey weights.

What is your longitudinal estimate of the union wage differential β? How does it differ from your cross-sectional estimate in question 1?

Answer:

```
. regress dlhw du [pweight = weight], robust
(sum of wgt is 227,467,244.8949)

Linear regression                               Number of obs     =     33,635
                                                F(1, 33633)       =      10.90
                                                Prob > F          =     0.0010
                                                R-squared         =     0.0006
                                                Root MSE          =     .38945

------------------------------------------------------------------------------
             |               Robust
        dlhw | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          du |   .0345348    .010462     3.30   0.001     .0140288    .0550407
       _cons |   .0383287   .0028503    13.45   0.000      .032742    .0439154
------------------------------------------------------------------------------
```

The point estimate $\widehat{\beta}^{\texttt{longi}} = 0.035$ is around a third of the cross-sectional estimate $\widehat{\beta}^{\texttt{cross}}$ from question 1, implying a longitudinal union wage differential of 3.5%

3. Notice that we can separate union membership changes across waves into two types of transition:

   (i) *union joiners*: $d_{i1}^{\texttt{union}} = 0$ and $d_{i5}^{\texttt{union}} = 1$

   (ii) *union leavers*: $d_{i1}^{\texttt{union}} = 1$ and $d_{i5}^{\texttt{union}} = 0$

   Estimate a *first-differenced* equation like (2) but decomposing transitions $\Delta d_i^{\texttt{union}}$ into joiners and leavers. That is, run regression

$$\Delta \log (w_i) = \gamma + \beta_1 \, d_i^{\texttt{join}} + \beta_2 \, d_i^{\texttt{leave}} + \Delta\varepsilon_i, \tag{3}$$

   where

$$d_i^{\texttt{join}} = \begin{cases} 1 & \text{if } \; d_{i1}^{\texttt{union}} = 0 \text{ and } d_{i5}^{\texttt{union}} = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad d_i^{\texttt{leave}} = \begin{cases} 1 & \text{if } \; d_{i1}^{\texttt{union}} = 1 \text{ and } d_{i5}^{\texttt{union}} = 0 \\ 0 & \text{otherwise} \end{cases} .$$

   Again, use the survey weights.

   What are your estimates of the effects of joining and leaving unions? How do they compare to your results from question 2?

   Answer:

```
. regress dlhw nu un [pweight = weight], robust
(sum of wgt is 227,467,244.8949)

Linear regression                               Number of obs     =     33,635
                                                F(2, 33632)       =       5.54
                                                Prob > F          =     0.0039
                                                R-squared         =     0.0006
                                                Root MSE          =     .38946
```

dlhw	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
nu	.0384568	.0144488	2.66	0.008	.0101367	.0667768
un	-.0299309	.0157944	-1.90	0.058	-.0608886	.0010267
_cons	.038013	.0029616	12.84	0.000	.0322082	.0438178

Point estimates $\widehat{\beta}_1^{\texttt{longi}} = 0.038$ and $\widehat{\beta}_2^{\texttt{longi}} = -0.03$ imply a wage increase of 3.8% when joining and a wage decrease of 3% when leaving a union, respectively. These estimates are close in magnitude to $\widehat{\beta}^{\texttt{longi}}$ from question 2.

4. Using your results from question 3, perform a statistical test for symmetry of the effects of leaving/joining a union, i.e., for $H_0 : \beta_1 = -\beta_2$. Interpret your finding.

Answer:

```
. test nu = -un

 ( 1)  nu + un = 0

       F(  1, 33632) =    0.15
            Prob > F =    0.6959
```

We cannot reject the null hypothesis that $\beta_1 = -\beta_2$, as can be seen from the large p-value $p = 0.696$. This is a test of the implicit linearity assumption in equation (2) that the effects of changing membership status from 0 to 1 and from 1 to 0 are the same and equal to β. One way to see this is by noting that equation (2) can be rewritten as

$$\Delta \log(w_i) = \gamma + \beta\, d_i^{\texttt{join}} - \beta\, d_i^{\texttt{leave}} + \Delta\varepsilon_i \tag{2'}$$

since

$$\Delta d_i^{\texttt{union}} = \begin{cases} 1 & \text{if} \quad (d_{i1}^{\texttt{union}}, d_{i5}^{\texttt{union}}) = (0,1) \\ 0 & \text{if} \quad (d_{i1}^{\texttt{union}}, d_{i5}^{\texttt{union}}) \in \{(0,0),(1,1)\} \\ -1 & \text{if} \quad (d_{i1}^{\texttt{union}}, d_{i5}^{\texttt{union}}) = (1,0) \end{cases}$$

$$= d_i^{\texttt{join}} - d_i^{\texttt{leave}}$$

$$\implies \beta\, \Delta d_i^{\texttt{union}} = \beta\, d_i^{\texttt{join}} - \beta\, d_i^{\texttt{leave}}.$$

So, equations (2) and (3) are equivalent under H_0.

5. Evaluate your results from questions 1–4 in light of the discussion about possible biases in cross-section and longitudinal estimates of the union wage differential. Briefly include a discussion of how misclassification of union status may affect this, in both the cross-section and longitudinal analysis.

Answer:

Cross-sectional estimates of the union wage differential most likely suffer from two main issues:

(i) self-selection/sorting into union jobs

(ii) misclassification of union status

Longitudinal estimates help us with (i) by removing any time-invariant individual effects. To fix ideas, suppose that $\varepsilon_{it} = \omega_i + \eta_{it}$ with η_{it} uncorrelated with union status and covariates x_i, but ω_i potentially correlated. First-differencing removes the ω_i unobserved individual effects so, to the extent that selection is on time-invariant unobservables, longitudinal estimates are free of selection bias.

However, under classical measurement error conditions, (ii) generates an attenuation bias in cross-sectional estimates that is exacerbated in longitudinal estimates. Therefore, it is not clear a priori whether the difference in estimates whereby $\widehat{\beta}^{\texttt{longi}} < \widehat{\beta}^{\texttt{cross}}$ stems from selection bias or attenuation bias due to measurement error. Also, notice that identification in the *first-differenced* specification comes from people moving in and out of union jobs, under the implicit assumption that these union status moves are random.