

## Econometrics II Assignment

This assignment is due on May 17 (Monday) 5pm. Please email your report to the following address: [peppeconometrics@gmail.com](mailto:peppeconometrics@gmail.com).

Please make the email subject and the file name as “Assignment + Name + Student number”. If you submit multiple files, please combine them into one. Late submission or emails to any other address will NOT be accepted.

- Two researchers consider the fixed effects regression model:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it},$$

where  $\alpha_i$  is the fixed effect and  $t = 1, 2, 3$  (three periods). They are concerned about the endogeneity of  $X_{it}$  and find an IV  $Z_i$  to deal with it. Unfortunately,  $Z_i$  is time-invariant (e.g., distance). Researcher A and B take the following two approaches respectively:

- First generate a time-varying instrument  $Z_{it} = Z_i \times t$  for  $t = 1, 2, 3$ . Then B uses  $Z_{it}$  as the instrument to obtain the panel 2SLS estimator of  $\beta_1$ . Denote it this estimator as  $\hat{\beta}_{1A}$ .
- Apply the first difference between period 1 and period 3 to obtain a cross sectional data. That is, B focuses on the following model:

$$Y_{i3} - Y_{i1} = \beta_1(X_{i3} - X_{i1}) + u_{i3} - u_{i1},$$

and uses  $Z_i$  to obtain the (cross-sectional) 2SLS estimator of  $\beta_1$ . Denote this estimator as  $\hat{\beta}_{1B}$ .

Compare  $\hat{\beta}_{1A}$  and  $\hat{\beta}_{1B}$ , are they reliable estimators? which one is better (or equally good)? Please explain the reason.

Hint: you may find the following two formulae useful:

- In a general fixed effect model with  $T$  periods and a time-varying IV  $Z_{it}$ , the 2SLS for the panel fixed effects regression model:

$$\tilde{\beta}_1 = \frac{\frac{1}{T} \sum_{t=1}^T \frac{1}{n} \sum_{i=1}^n (Y_{it}^* Z_{it})}{\frac{1}{T} \sum_{t=1}^T \frac{1}{n} \sum_{i=1}^n (X_{it}^* Z_{it})}, \quad (1)$$

where  $X_{it}^* = X_{it} - \bar{X}_i$ ,  $Y_{it}^* = Y_{it} - \bar{Y}_i$ ,  $u_{it}^* = u_{it} - \bar{u}_i$ , and  $\bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it}$ ,  $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}$ ,  $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$ .

- The 2SLS estimator for cross-sectional data set  $\{(Y_i, X_i, Z_i)\}_{i=1}^n$  can be written as

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum_{i=1}^n Y_i Z_i}{\frac{1}{n} \sum_{i=1}^n X_i Z_i}. \quad (2)$$

2. Let  $Z_1$  and  $Z_2$  be continuous variables,  $D_1$  be a dummy variable. Consider the following probit model:

$$\Pr(Y = 1 \mid Z_1 = z_1, Z_2 = z_2, D_1 = d_1) = \Phi(\delta_1 z_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1).$$

- (a) Find the partial effect of  $Z_2$
- (b) Find the effect of  $D_1$  on the probability.
- (c) How would you estimate these two effects?
3. You are asked to reproduce some of the results of the paper. But don't worry if your results are not exactly the same as those in the paper.

*Card, David and Alan B. Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania" American Economic Review, Vol. 84 (September 1994): 772-793.* The data used by Card and Krueger is provided under the name: **fastfood.dta**. Refer to the article for explanations of the variables. You need to read the paper carefully in order to answer the following questions. **Parts of the Stata code are provided in brackets.**

- (a) Suppose the research wants to use Differences-in-differences (DID) strategy to study the disemployment effect of the rise of minimum wage. What is the main assumption for DID?
- (b) Taking into account another group of employees (in both states NJ and PA) who received a higher wage (over 5.05 dollar per hour) before the rise of minimum wage in April 1992, the researcher may use Differences-in-difference-in-differences (DDD) strategy. Let us construct a variable  $L = 1$  if the employee received low wage initially (before April 1992), and  $L = 0$  if the employee received high wage initially. Fully write out the regression equation for DDD, and indicate which coefficient measures the treatment effect. (For simplicity, you can ignore the covariates.)
- (c) First confirm that your data is similar to that used by authors, by reproducing panel 1. of Table 2.

(tab chain state, column)

Next, construct a first measure of change in FTE (full time equivalent) employment similar to the one used in the article with the following commands

```

(gen fte1= empft+nmgrs+0.5*emppt
gen fte2= empft2+nmgrs2+0.5*emppt2
gen dfte=dfte2-fte1
gen NJ=1
replace NJ=0 if pa1==1 | pa2==1
regress dfte NJ, robust)

```

Note that column (iii), rows 3, 4, and 5 of Table 3 of the paper used different measures of dfte. Please do the same thing by modifying the sample restrictions and the definition of dfte (with different measures of employment or subset of stores). Does the coefficient of NJ match the estimates presented in column (iii), rows 3,

4, and 5 of Table 3 of the paper? Construct a Table 1 with 4 columns that shows these results and indicate how many stores are included at the bottom of each column. To be specific, Column (1) does not impose additional restrictions on the sample. Column (2) restricts the sample to the stores that are permanently open in both periods. Column (3) puts 0 on the employment for those temporarily opened stores. Column (4) restricts the sample to that have wage data available in both periods.

- (d) In equation (1a) of the paper, the effect of the New Jersey dummy was estimated with a differenced equation:

$$\Delta E_i = a + bX_i + cNJ + \varepsilon_i \quad (3)$$

Now we try an alternative approach. Use the following commands to expand the dataset. (They transform the data from a wide format to a long format. You can experiment with the code to see how they work. The variable T is generated to indicate period 1 or 2.)

```
(expand 2
sort sheet
by sheet: gen T = _n - 1
gen TNJ=T*NJ
gen fte=fte1 if T==0
replace fte=fte2 if T==1)
```

Now estimate explicitly a difference-in-differences model with explanatory variables,

$$E_i = \alpha + \beta X_i + \delta_t T + \delta_{NJ} NJ + \gamma TNJ + \varepsilon_i, \quad (4)$$

using fte as the dependent variable and T, NJ and TNJ as the regressors.

[1]. Which coefficient corresponds to the treatment effect?

[2]. Do you obtain the same results as in (3) for the same subset of stores? (You do not need to estimate (3), as the results are given in column (i) and (ii) in Table 4 of the paper.)

[3]. Report your results in a Table 2. You can use multiple columns with different measures of fte employment or/and with different specifications (for example, you can include dummies for regions). Summarize your findings in several sentences.

[4]. Explicitly state the conditions (or scenarios) under which regressions (3) and (4) yield the same estimates.

- (e) Given that no disemployment effects on minimum wage workers of the increase in the minimum wage are found, Card and Krueger consider many other possible adjustments of the fast-food chains. Another possibility is negative employment effects on higher paid workers. Perform the estimation in the regression equation

(3) using (as the dependent variable) the number of managers in two periods (*nmgrs* and *nmgrs2*) to compute changes in employment. (Here you need to reload the original data set, instead of using the transformed one in the previous part.) Report your estimates in a Table 3 and summarize your findings in several sentences.