

Problem Set V

Econometrics III

Nurfatima Jandarova

March 7, 2017

1

1.1 By including $d2_t$ we make sure to capture the common factors that could have been affecting y_{it} even in the absence of the treatment. In a sense, this makes sure we capture the aggregate trend in the economy (e.g., different stages of the business cycle at two periods of the panel data) that is assumed to be common across all cross-section units. Omitting $d2_t$ would make the estimate of the treatment effect biased in the same direction as the common trend. For example, if there was a general upward trend in the economy, then the estimate of the treatment effect in absence of $d2_t$ would have also picked up the time trend and thus be overestimated.

1.2 Since nothing is said about the procedure for assignment of treatment, I would say that the main reason for including c_i , possibly unobservable, is to capture any potential non-randomness of the treatment assignment. Then, by either taking first differences (as below) or using fixed effects estimation, one could take care of the non-randomness such that the estimate of δ_1 could be interpreted causally.

1.3 Define the following objects:

$$\begin{aligned} N_c &= \sum_{i=1}^N \mathbb{1}\{prog_{i2} = 0\} \\ N_t &= N - N_c \\ \bar{\Delta}y_{control} &= \frac{1}{N_c} \sum_{i=1}^N [\mathbb{1}\{prog_{i2} = 0\} \Delta y_i] \\ \bar{\Delta}y_{treat} &= \frac{1}{N_t} \sum_{i=1}^N [\mathbb{1}\{prog_{i2} = 1\} \Delta y_i] \end{aligned}$$

Rewrite the model in first differences:

$$\Delta y_i = \theta_2 + \delta_1 prog_{i2} + \Delta u_i$$

Notice that using the assumption in the problem $\mathbb{E}(u_{it}|prog_{i2}, c_i) = 0$, we can also write $\mathbb{E}(\Delta u_i|prog_{i2}) = \mathbb{E}[\mathbb{E}(\Delta u_i|prog_{i2}, c_i)] = 0$. Hence, one could find the estimators $\hat{\theta}_2$ and $\hat{\delta}_1$ by minimizing $\sum_{i=1}^N (\Delta \hat{u}_i)^2$. To simplify computation slightly, I first consider subsample for which $prog_{i2} = 0$:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{prog_{i2} = 0\} \hat{\theta}_2 &= \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{prog_{i2} = 0\} \Delta y_i \\ \hat{\theta}_2 &= \frac{1}{N_c} \sum_{i=1}^N [\mathbb{1}\{prog_{i2} = 0\} \Delta y_i] = \bar{\Delta}y_{control} \end{aligned}$$

Using this result, similarly for a subsample for which $prog_{i2} = 1$:

$$\begin{aligned}\frac{1}{N} \sum_{i=1}^N \left[\mathbb{1}\{prog_{i2} = 1\} (\Delta y_i - \hat{\theta}_2 - \hat{\delta}_1) \right] &= 0 \\ N_t \bar{\Delta} y_{treat} - N_t \bar{\Delta} y_{control} - N_t \hat{\delta}_1 &= 0 \\ \bar{\Delta} y_{treat} - \bar{\Delta} y_{control} &= \hat{\delta}_1\end{aligned}$$

1.4 A simplest extension of the model would be to include time dummies for all additional time periods (except first since we have a constant):

$$y_{it} = \theta_1 + \theta_2 dk_{2t} + \dots + \theta_T dk_{Tt} + \delta_1 prog_{it} + c_i + u_{it}$$

where as before $dk_t = \mathbb{1}\{t = k\}, \forall k \in \{2, \dots, T\}$.

1.5 I would prefer the extension from section 1.4 because it takes into account "new information" available at time $t > 2$. One could also extend the model to allow for non-linearity in the treatment effects. On the other hand, the equation specified in the question only uses information available at time $t = 1$ and $t = 2$; hence, is inefficient. On top of that, it also restricts us to look at linear treatment effects, so another dimension in which we would be losing information.

2

We are given the following model to estimate

$$\ln(wage_{it}) = \alpha + \beta_1 educ_i + \beta_2 black_i + \beta_3 hisp_i + \beta_4 exper_{it} + \beta_5 exper_{it}^2 + \beta_6 married_{it} + \beta_7 union_{it} + \theta_t + c_i + \varepsilon_{it}$$

Since we are not given any assumption about the nature of time fixed effects, I could either introduce time dummies to stand for θ_t or a time trend (linear, for simplicity). The first option is great whenever we are not sure about the nature of time trend, and we might not be comfortable enough to assume linearity. But with $T > 2$, the time dummies reduce degrees of freedom, especially if one wants to interact regressors with time fixed effects. On the other hand, assuming a simple linear time trend may save us lots of trouble with degrees of freedom. Moreover, the sample years 1980 - 1987 seem to be safe enough to assume linearity as the time is usually referred to as Great Moderation. Therefore, in the following I will assume that the time trend is linear¹.

2.3 The pooled OLS estimation results are presented in columns (1) and (2) of Table 1. Note that we cannot rely on usual OLS standard errors because of serial correlation. Even if θ_t and ε_{it} are serially uncorrelated, the presence of unobserved heterogeneity (c_i) introduces serial correlation of estimated error term. To see this, consider an individual i at time t and $s \neq t$:

$$\mathbb{E}[(\theta_t + c_i + \varepsilon_{it})(\theta_s + c_i + \varepsilon_{is})] = \mathbb{E}[c_i^2] = \sigma_c^2 > 0$$

Hence, the estimated residuals are serially correlated. To correct for this serial correlation, one could compute clustered standard errors on the individuals (presented in column (2) of Table 1). One may observe that clustered errors are higher than usual standard errors, which points to a positive intracluster correlations.

2.4 The random effect estimates are presented in column (3) of Table 1. The point estimates in this case are relatively similar to pooled OLS coefficients, while standard errors are typically smaller compared to clustered pooled OLS, as one might have anticipated. The differences in point estimates could be attributed to the transformation of data (weighting) involved in estimating random effects, although both methods should yield consistent estimates under assumption of strict exogeneity. However, the difference

¹I also done estimations using the dummy variables and qualitatively results seem not to change much. Although some of the coefficients do differ between the two specifications. Also, I find it a bit puzzling that the coefficients for the time dummies change considerably between RE and FE specifications, while in my opinion they shouldn't. All that FE does to the dummies is it subtracts a constant $\frac{1}{T}$ from all of them, so it should affect α , but not the coefficient on the dummy.

²Assume $c_i \sim \text{i.i.d. } \mathcal{N}(0, \sigma_c^2)$ and that c_i is uncorrelated with ε_{it} and θ_t .

	(1) POLS standard	(2) POLS clustered	(3) RE	(4) FE	(5) FE	(6) FE	(7) FE
exper	0.065*** (0.012)	0.065*** (0.017)	0.094*** (0.014)	0.117*** (0.008)	0.134*** (0.024)	0.170*** (0.027)	0.103*** (0.013)
expersq	-0.002*** (0.001)	-0.002*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.004*** (0.001)
married	0.108*** (0.016)	0.108*** (0.026)	0.063*** (0.017)	0.045** (0.018)	0.045** (0.018)	0.047*** (0.018)	0.049** (0.022)
union	0.183*** (0.017)	0.183*** (0.027)	0.108*** (0.018)	0.082*** (0.019)	0.082*** (0.019)	0.079*** (0.019)	0.064*** (0.023)
1980b.year#c.educ						0.000 (0.000)	
1981.year#c.educ						-0.001 (0.003)	
1982.year#c.educ						-0.006 (0.004)	
1983.year#c.educ						-0.011** (0.006)	
1984.year#c.educ						-0.014* (0.007)	
1985.year#c.educ						-0.016* (0.009)	
1986.year#c.educ						-0.017* (0.010)	
1987.year#c.educ						-0.017 (0.011)	
educ	0.091*** (0.005)	0.091*** (0.011)	0.092*** (0.011)				
black	-0.139*** (0.024)	-0.139*** (0.050)	-0.139*** (0.048)				
hisp	0.016 (0.021)	0.016 (0.039)	0.021 (0.043)				
year	0.022*** (0.006)	0.022* (0.012)	0.017 (0.011)				
c.educ#c.year					-0.001 (0.002)		
F.union							0.071*** (0.023)
Constant	-43.158*** (12.789)	-43.158* (23.423)	-34.073 (22.516)	1.065*** (0.027)	28.357 (35.619)	0.920*** (0.077)	1.105*** (0.042)
Observations	4,360	4,360	4,360	4,360	4,360	4,360	3,270
R-squared	0.189	0.189		0.178	0.178	0.181	0.108
Number of nr			545	545	545	545	545

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Table 1: Estimation results

between some coefficients (e.g., on *married* and *union*) is beyond the 95% confidence interval³. This could in fact suggest correlation between the individual-specific effect and the regressors, which violates random effects assumption.

2.5 I have *year* to drop out of my estimation because it comes after *exper*, just a technical note. But the reason is basically that after within transformation the two variables become identical, so there is a problem of perfect collinearity and we have to drop one of them.

Another "interesting"⁴ thing to notice is that *educ_{it}* becomes a redundant variable. Since in our model *educ* is individual-specific and time-invariant, it got differenced out by within transformation just as *black* and *hisp*. Comparing the coefficients in columns (3) and (4) of Table 1, we could see that FE point estimates of β_6 and β_7 are lower than RE, while β_4 seems to be considerably larger. This again suggests that *married* and *union* were correlated with individual-specific effects. Hence, now that we got rid of individual-specific effects, we also took care of the bias it created.

2.6 To follow up the guess from previous sections, Hausman test allows us to test if the strict exogeneity assumption of RE is violated or not. The test statistic $\chi^2_4 = 16.64$ and p-value = 0.0023. Hence, the Hausman test of fixed versus random effects suggests that the two sets of coefficients are different and that FE is preferred to RE.

2.7 The first-differenced estimation results are presented in Table 2. The coefficients seem to be fairly close to the ones obtained by fixed effects estimation. The coefficient on $\Delta union$ is "lower" than in FE specification; however, the difference between the two fits into $\sim 2sd$. Still this might suggest that strict exogeneity assumption for *union* fails.

VARIABLES	(1) D.lwage
D.exper	0.116*** (0.020)
D.expersq	-0.004*** (0.001)
D.married	0.038* (0.023)
D.union	0.043** (0.020)
Observations	3,815
R-squared	0.027
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Table 2: Estimation results: FD

2.8 I used *xtserial* command that "implements a test for serial correlation in the idiosyncratic errors of a linear panel-data model discussed by Wooldridge (2002)". According to the test results below, there is an *AR*(1) serial correlation in the error term of the first differenced regression.

```
H0: no first order autocorrelation
F( 1,      544) =      25.372
Prob > F =      0.0000
```

³I tried to perform Hausman test using the two regressions, but as it turns out the test is "bad" if one suspects clustered standard errors as is the case here. Unfortunately, I do not yet know a more sophisticated way to test the equality of coefficients in two regressions.

⁴It was sort of interesting before I realized I forgot to include time specific effects and the question why *exper* drops out did not make sense to me.

2.9 To see if the education premium somehow changed over time we need to add interaction term in our regression (column (5) of Table 1). The results suggest that there is no statistically significant difference in returns to education over time. However, if I include the interaction of education level with time dummies, I get that education premium was about 1.14 - 1.70pp lower in 1983 - 1986 than in 1980, considerable and statistically significant effect.

2.10 The result of FE estimation with one-period lead of *union* is presented in column (7) of Table 1. It suggests that the one-period lead of *union* is statistically significant. This indeed suggests that strict exogeneity assumption on *union* fails.

3

3.1 The pooled OLS specification estimates the following regression model:

$$d_{it} = \alpha d_{it-1} + \gamma y_{it-1} + \mathbf{x}'_{it-1} \beta + \mu_t + v_{it}$$

i.e., where $v_{it} = \delta_i + u_{it}$. In this case, pooled OLS estimates are inconsistent because exogeneity assumption necessarily fails due to lagged dependent variable, $\mathbb{E}(d_{it-1} v_{it}) = \mathbb{E}(d_{it-1} (\delta_i + u_{it})) \neq 0^5$. Although, the main argument by Acemoglu et al. (2008) is that OLS is inconsistent whenever $Cov(\mathbf{x}'_{it-1}, \delta_i + u_{it}) \neq 0$ or $Cov(y'_{it-1}, \delta_i + u_{it}) \neq 0$. However, I do not quite agree because already the presence of individual-specific effects creates bias.

3.2 After first differencing the model, one gets

$$\Delta d_{it} = \alpha \Delta d_{it-1} + \gamma \Delta y_{it-1} + \Delta \mathbf{x}'_{it-1} \beta + \Delta \mu_t + \Delta u_{it}$$

If the original error term u_{it} is stationary, then Δu_{it} would be serially correlated. Hence, the above model cannot be estimated consistently. However, one could use d_{it-2} as an instrument for Δd_{it-1} . First, the instrument is relevant because $Cov(d_{it-2}, \Delta d_{it-1}) = Cov(d_{it-2}, d_{it-1} - d_{it-2}) = \alpha + Var(d_{it-2}) \neq 0$. Second, the instrument is exogenous

$$\mathbb{E}(d_{it-2}, \Delta u_{it}) = \mathbb{E}(\alpha d_{it-3} + \gamma y_{it-3} + \mathbf{x}'_{it-3} \beta + \mu_{t-2} + \delta_i + u_{it-2}, u_{it} - u_{it-1}) = 0$$

under assumption that past values of regressors are not correlated with future shocks.

3.3 Acemoglu et al. (2008) mention that "under the assumption of no further serial correlation", i.e., the true model indeed only has one-period lag of the dependent variable, not only d_{it-2} could be used as a valid instrument, but any $d_{i\tau}, \forall \tau \leq t-2$. The intuitive argument is that by not using the rest of the lagged values, we are not using full information the sample contains. So, by taking them into account, one can significantly improve efficiency of the estimators. As to the mechanics of the process, I hope to get better understanding during the exercise class.

⁵With the true model as given in (1) in Acemoglu et al. (2008), the lagged dependent variable would also be a function of δ_i ; thus creating a bias.