International School of Economics at TSU
Econometrics 2
Lasha Chochua

Problem Set 9

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Consider the binary variable version of the fixed effects model discussed in the class except with an additional regressor, $D1_i$; that is, let

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2_i + \cdots + \gamma_n Dn_i + u_{it}.$$

a. Suppose that $n = 3$. Show that the binary regressors and the "constant" regressor are perfectly multicollinear; that is, express one of the variables $D1_i, D2_i, D3_i$, and $X_{0,it}$ as a perfect linear function of the others, where $X_{0,it} = 1$ for all i, t.

b. Show the result in (a) for general n.

c. What will happen if you try to estimate the coefficients of the regression by OLS?

Problem 2: A researcher believes that traffic fatalities increase when roads are icy and thinks that therefore states with more snow will have more fatalities than other states. Comment on the following methods designed to estimate the effect of snow on fatalities:

a. The researcher collects data on the average snowfall for each state and adds this regressor ($AverageSnow_i$) to the regressions given in Table 10.1.

b. The researcher collects data on the snowfall in each state for each year in the sample and adds this regressor to the regressions.

Problem 3:

a. In the fixed effects regression model, are the fixed entity effects, α_i, consistently estimated as $n \longrightarrow \infty$ with T fixed?

b. If n is large (say, $n = 2000$) but T is small (say, $T = 4$), do you think that the estimated values of α_i are approximately normally distributed? Why or why not?

Problem 4: Let $\hat{\beta}_1^{DM}$ denote the entity-demeaned estimator given in Equation (10.22), and let $\hat{\beta}_1^{BA}$ denote the "before and after" estimator without an intercept, so that

$$\hat{\beta}_1^{BA} = \left[\sum_{i=1}^{n}(X_{i2} - X_{i1})(Y_{i2} - Y_{i1})\right] \Big/ \left[\sum_{i=1}^{n}(X_{i2} - X_{i1})^2\right].$$

Show that, if $T = 2$, $\hat{\beta}_1^{DM} = \hat{\beta}_1^{BA}$.

Problem 5: You wish to study the effects of unionisation on wages using a panel of N individuals and T time periods. You wish to allow for the following phenomena:

- (a) Unionised firms select the higher ability workers
- (b) Workers with bad productivity shocks join the union sector

a. Set up a suitable model and explain how these phenomena are reflected in your specification.

b. Explain how you would estimate this model and present the estimator. Carefully state any assumptions you make.

Problem 6: A common setup for program evaluation with two periods of panel data is the following. Let y_{it} denote the outcome of interest for unit i in period t.

At $t = 1$, no one is in the program. At $t = 2$, some units are in the control group and others are in the experimental group. Let $prog_{it}$ be a binary indicator equal to one if unit i is in the program in period t; by design, $prog_{i1} = 0$ for all i.

An unobserved effects model without additional covariates is:

$$y_{it} = \theta_1 + \theta_2 d2_t + \delta_1 prog_{it} + c_i + u_{it}, \qquad \mathbb{E}(u_{it} \mid prog_{i2}, c_i) = 0$$

where $d2_t = 1$ if $t = 2$, and 0 otherwise; c_i is the unobserved individual effect.

a. Explain why including $d2_t$ is important in these contexts. In particular, what problems might be caused by leaving it out?

b. Why is it important to include c_i in the equation?

c. Using the first differencing method, show that:

$$\hat{\theta}_2 = \overline{\Delta y}_{control}, \quad \hat{\delta}_1 = \overline{\Delta y}_{treat} - \overline{\Delta y}_{control}$$

where $\overline{\Delta y}_{control}$ is the average change in y for units with $prog_{i2} = 0$, and $\overline{\Delta y}_{treat}$ is the average change for units with $prog_{i2} = 1$. This shows that $\hat{\delta}_1$, the **difference-in-differences estimator**, arises from an unobserved effects panel model.

d. Write down the extension of the model for T time periods.

Problem 7: Consider the model:

$$y = z_1\beta + w\alpha + \varepsilon$$

where:

- $E(z\varepsilon) = 0$
- $z = (z_1, z_2)$ is a vector of exogenous variables
- w is endogenous: $E(w\varepsilon) \neq 0$

Suppose we estimate (β, α) using the following two-step procedure:

Step 1: Regress w on z_2 and obtain the fitted values $\hat{w}$.
Step 2: Regress y on $(z_1, \hat{w})$ and obtain $(\hat{\beta}, \hat{\alpha})$.

a. Will $(\hat{\beta}, \hat{\alpha})$ be generally consistent? Show.

b. When will $(\hat{\beta}, \hat{\alpha})$ be consistent?

Problem 8: Suppose you wish to estimate β in:

$$y_i = \alpha + x_i\beta + u_i$$

a. Derive the consequences for the OLS estimator if y is measured with error that is independent of x.

b. Instead of measuring x you measure x^* where $x_i^* = x_i + \epsilon_i$ and ϵ_i is a measurement error which is independent across individuals and independent of x. Show that the OLS estimator converges asymptotically to $\delta\beta$ where $0 \leq \delta \leq 1$. Explain the implication of this result for estimating the elasticity of hours worked with respect to wages when wages are measured with iid errors.

If x is measured with error, so that we observe x^* where

$$x_i^* = x_i + \epsilon_i$$

where x is independent of ϵ.