International School of Economics at TSU
Econometrics 2
Lasha Chochua

Problem Set 11

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: Efficiency of Differences vs. Differences-in-Differences

Suppose there are panel data for $T = 2$ time periods for a randomized controlled experiment. The first observation ($t = 1$) is taken before the experiment and the second ($t = 2$) is the post-treatment period. The treatment is binary: $X_{it} = 1$ if individual i is in the treatment group and $t = 2$, and $X_{it} = 0$ otherwise. The outcome follows

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

where α_i are individual-specific effects with $E[\alpha_i] = 0$ and $\text{Var}(\alpha_i) = \sigma_\alpha^2$; the error u_{it} is homoskedastic with variance σ_u^2, $\text{cov}(u_{i1}, u_{i2}) = 0$, and $\text{cov}(u_{it}, \alpha_i) = 0$ for all i. Denote by $\hat{\beta}_1^{\text{diff}}$ the OLS estimator from regressing Y_{i2} on X_{i2} with an intercept, and by $\hat{\beta}_1^{\text{DiD}}$ the OLS estimator from regressing $\Delta Y_i = Y_{i2} - Y_{i1}$ on $\Delta X_i = X_{i2} - X_{i1}$ with an intercept.

a. Show that $n \cdot \text{Var}(\hat{\beta}_1^{\text{diff}}) \to (\sigma_\alpha^2 + \sigma_u^2) \,/\, \text{Var}(X_{i2})$.

b. Show that $n \cdot \text{Var}(\hat{\beta}_1^{\text{DiD}}) \to 2\sigma_u^2 \,/\, \text{Var}(X_{i2})$. (*Hint:* First explain why $X_{i2} - X_{i1} = X_{i2}$.)

c. Based on (a) and (b), under what condition is $\hat{\beta}_1^{\text{DiD}}$ more efficient than $\hat{\beta}_1^{\text{diff}}$? Interpret the condition economically.

Problem 2: DiD as a Levels Regression

You have panel data with $T = 2$ periods. Let $G_i \in \{0, 1\}$ indicate the treatment group and $D_t \in \{0, 1\}$ indicate the second period ($t = 2$). The treatment indicator is $X_{it} = G_i \cdot D_t$. Consider the pooled OLS regression

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 G_i + \beta_3 D_t + u_{it}.$$

Show that the OLS estimator $\hat{\beta}_1$ equals the differences-in-differences estimator

$$\widehat{\beta}_1^{\text{DiD}} = (\bar{Y}_{\text{treat,after}} - \bar{Y}_{\text{treat,before}}) - (\bar{Y}_{\text{control,after}} - \bar{Y}_{\text{control,before}}).$$

Problem 3: OLS with Heterogeneous Treatment Effects

Consider the random coefficients model

$$Y_i = \beta_0 + \beta_{1i} X_i + v_i,$$

where (v_i, X_i, β_{1i}) are i.i.d., $X_i \in \{0, 1\}$ is a binary treatment, and $\beta_1 = E[\beta_{1i}]$.

a. Show that the model can be written as $Y_i = \beta_0 + \beta_1 X_i + u_i$, defining u_i explicitly.

b. Suppose X_i is randomly assigned so that $X_i \perp (\beta_{1i}, v_i)$. Show that $E[u_i \mid X_i] = 0$.

c. Hence show that $\hat{\beta}_1 \xrightarrow{p} E[\beta_{1i}]$, i.e., OLS consistently estimates the average treatment effect.

d. Verify that Assumptions 1 and 2 of the large-sample OLS key concept are satisfied.

e. Now suppose X_i is *not* randomly assigned, that $E[v_i \mid X_i] = 0$, but that $\text{cov}(\beta_{1i}, X_i) > 0$ – individuals with larger-than-average treatment gains are more likely to be treated. Which OLS assumption fails? Is $\hat{\beta}_1$ consistent for $E[\beta_{1i}]$? Find the direction of the bias.

Problem 4: LATE, Compliers, and the Limits of IV

A government randomly assigns a lottery indicator $Z_i \in \{0, 1\}$ to n workers ($Z_i = 1$ means selected). Workers then choose whether to enroll in a job-training program ($X_i = 1$). Assume **monotonicity** (no defiers). The outcome Y_i is annual wages in USD. The following quantities are observed from a large sample:

$$P(X_i = 1 \mid Z_i = 1) = 0.6, \qquad P(X_i = 1 \mid Z_i = 0) = 0.2,$$

$$E[Y_i \mid Z_i = 1] = 51{,}000, \qquad E[Y_i \mid Z_i = 0] = 48{,}200.$$

a. Under monotonicity, define the three types of individuals (always-takers, compliers, never-takers) and compute the proportion of each type in the population.

b. Compute the TSLS (Wald) estimator. Show your work.

c. Using the framework of Section 13.6, identify the causal parameter estimated by TSLS. Which subpopulation drives the estimate?

d. Suppose you additionally learn that the average treatment effects by type are:

$$\begin{aligned} E[Y_i(1) - Y_i(0) \mid \text{always-taker}] &= 3{,}000, \\ E[Y_i(1) - Y_i(0) \mid \text{complier}] &= 7{,}000, \\ E[Y_i(1) - Y_i(0) \mid \text{never-taker}] &= 1{,}000. \end{aligned}$$

Compute the population ATE. Verify numerically that the Wald formula in (b) recovers exactly the complier average effect, not the ATE.

e. Explain intuitively why TSLS cannot recover the ATE in this setting, even though Z_i is randomly assigned and the instrument is valid.

Problem 5: What Can We Learn from a Randomized Experiment?

Let $X_i \in \{0,1\}$ be a binary treatment, **randomly assigned** and therefore independent of $(Y_i(0), Y_i(1))$. Define the individual treatment effect $\text{TE}_i = Y_i(1) - Y_i(0)$, and write $\sigma^2_{Y(d)} = \text{Var}(Y_i(d))$ for $d \in \{0,1\}$.

a. Show that $E[Y_i(1)]$ and $E[Y_i(0)]$ are consistently estimable. Provide explicit estimators.

b. Show that $E[\text{TE}_i]$ is consistently estimable, and that the differences estimator $\bar{Y}_{\text{treat}} - \bar{Y}_{\text{control}}$ is consistent for it.

c. Show that $\text{Var}(Y_i(1))$ and $\text{Var}(Y_i(0))$ are consistently estimable. Propose consistent estimators.

d. Show that

$$\text{Var}(\text{TE}_i) = \sigma^2_{Y(1)} + \sigma^2_{Y(0)} - 2\,\text{Cov}(Y_i(1),\, Y_i(0)).$$

Hence argue that $\text{Var}(\text{TE}_i)$ is **not** consistently estimable from a single randomized experiment, even with infinite data. What is the unidentified object?

e. Despite (d), show that the following sharp bounds hold:

$$\left(\sigma_{Y(1)} - \sigma_{Y(0)}\right)^2 \;\le\; \text{Var}(\text{TE}_i) \;\le\; \left(\sigma_{Y(1)} + \sigma_{Y(0)}\right)^2.$$

Both bounds are consistently estimable. Interpret when each bound is achieved.