**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 $\alpha_i$ are individual-specific effects with $E[\alpha_i] = 0$ and $\mathrm{Var}(\alpha_i) = \sigma_\alpha^2$; the error $u_{it}$ is homoskedastic with variance $\sigma_u^2$, $\mathrm{cov}(u_{i1}, u_{i2}) = 0$, and $\mathrm{cov}(u_{it}, \alpha_i) = 0$ 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 $n \cdot \mathrm{Var}(\hat{\beta}_1^{\text{diff}}) \to (\sigma_\alpha^2 + \sigma_u^2) \,/\, \mathrm{Var}(X_{i2})$.

**b.** Show that $n \cdot \mathrm{Var}(\hat{\beta}_1^{\text{DiD}}) \to 2\sigma_u^2 \,/\, \mathrm{Var}(X_{i2})$. (*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.

**Solution:**

**a.** In period 2 the model gives $Y_{i2} = \alpha_i + \beta_1 X_{i2} + u_{i2}$. The differences estimator regresses $Y_{i2}$ on $X_{i2}$ treating the composite $\varepsilon_i = \alpha_i + u_{i2}$ as the regression error. By the homoskedasticity-only variance formula (Appendix 5.1),

$$\mathrm{Var}(\hat{\beta}_1^{\text{diff}}) = \frac{\mathrm{Var}(\varepsilon_i)}{\sum_{i=1}^{n} (X_{i2} - \bar{X}_2)^2}.$$

Since $\text{cov}(\alpha_i, u_{i2}) = 0$, we have $\text{Var}(\varepsilon_i) = \sigma_\alpha^2 + \sigma_u^2$. Multiplying by $n$ and using $n^{-1}\sum_i(X_{i2} - \bar{X}_2)^2 \xrightarrow{p} \text{Var}(X_{i2})$,

$$\boxed{n \cdot \text{Var}(\hat{\beta}_1^{\text{diff}}) \;\to\; \frac{\sigma_\alpha^2 + \sigma_u^2}{\text{Var}(X_{i2})}}.$$

**b.** Because $X_{it} = 1$ only if the individual is treated *and* $t = 2$, we have $X_{i1} = 0$ for everyone in both groups: no individual is treated in period 1. Therefore $\Delta X_i = X_{i2} - X_{i1} = X_{i2}$.

Taking first differences of the model:

$$\Delta Y_i = Y_{i2} - Y_{i1} = (\alpha_i + \beta_1 X_{i2} + u_{i2}) - (\alpha_i + 0 + u_{i1}) = \beta_1 X_{i2} + (u_{i2} - u_{i1}).$$

The individual effect $\alpha_i$ cancels exactly. The DiD estimator regresses $\Delta Y_i$ on $\Delta X_i = X_{i2}$, so its error is $\eta_i = u_{i2} - u_{i1}$ with

$$\text{Var}(\eta_i) = \text{Var}(u_{i2}) + \text{Var}(u_{i1}) - 2\,\text{cov}(u_{i1}, u_{i2}) = \sigma_u^2 + \sigma_u^2 - 0 = 2\sigma_u^2.$$

Applying the same homoskedasticity-only formula:

$$\boxed{n \cdot \text{Var}(\hat{\beta}_1^{\text{DiD}}) \;\to\; \frac{2\sigma_u^2}{\text{Var}(X_{i2})}}.$$

**c.** $\hat{\beta}_1^{\text{DiD}}$ is more efficient when its asymptotic variance is smaller:

$$\frac{2\sigma_u^2}{\text{Var}(X_{i2})} < \frac{\sigma_\alpha^2 + \sigma_u^2}{\text{Var}(X_{i2})} \iff \boxed{\sigma_\alpha^2 > \sigma_u^2}.$$

Differencing eliminates $\alpha_i$ at the cost of adding $u_{i1}$ noise to the error. This trade is worthwhile whenever unobserved individual heterogeneity ($\sigma_\alpha^2$) is the larger source of outcome variation. In panel surveys, individual-level fixed factors – baseline health, ability, family background – are typically large relative to period-specific shocks, so DiD is usually preferred on efficiency grounds.

### 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

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

**Solution:**

The four regressors are $(1,\ G_iD_t,\ G_i,\ D_t)$. The covariate values for the four distinct $(G, D)$ cells are:

| $G$ | $D$ | 1 | $G_iD_t$ | $G_i$ | $D_t$ |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 |

This $4 \times 4$ design matrix has full rank (its determinant is non-zero), so the model is **saturated**: it has exactly as many free parameters as there are cells, and OLS fits each cell mean exactly. Denoting $\bar{y}_{gd}$ as the sample mean of $Y_{it}$ in cell $(G_i = g, D_t = d)$, OLS must satisfy:

$$\begin{aligned}
(0,0): &\quad \bar{y}_{00} = \hat{\beta}_0 \\
(0,1): &\quad \bar{y}_{01} = \hat{\beta}_0 + \hat{\beta}_3 \\
(1,0): &\quad \bar{y}_{10} = \hat{\beta}_0 + \hat{\beta}_2 \\
(1,1): &\quad \bar{y}_{11} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3.
\end{aligned}$$

The first three equations give $\hat{\beta}_0 = \bar{y}_{00}$, $\hat{\beta}_3 = \bar{y}_{01} - \bar{y}_{00}$, and $\hat{\beta}_2 = \bar{y}_{10} - \bar{y}_{00}$. Substituting into the fourth:

$$\hat{\beta}_1 = \bar{y}_{11} - \bar{y}_{00} - (\bar{y}_{10} - \bar{y}_{00}) - (\bar{y}_{01} - \bar{y}_{00}) = \bar{y}_{11} - \bar{y}_{10} - \bar{y}_{01} + \bar{y}_{00}.$$

Identifying the cell means with the group-time averages:

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

This result holds for any sample sizes within each cell (not only balanced designs), because the saturated model always interpolates cell means exactly.

**Problem 3: OLS with Heterogeneous Treatment Effects**

Consider the random coefficients model

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

where $(v_i, X_i, \beta_{1i})$ 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.

**Solution:**

**a.** Add and subtract $\beta_1 X_i$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_{1i} X_i + v_i \\ &= \beta_0 + \beta_1 X_i + (\beta_{1i} - \beta_1) X_i + v_i \\ &\equiv \beta_0 + \beta_1 X_i + u_i, \end{aligned}$$

where $u_i = (\beta_{1i} - \beta_1) X_i + v_i$. $\square$

**b.** Taking the conditional expectation of $u_i$:

$$\begin{aligned} E[u_i \mid X_i] &= E[(\beta_{1i} - \beta_1) X_i + v_i \mid X_i] \\ &= X_i \cdot E[\beta_{1i} - \beta_1 \mid X_i] + E[v_i \mid X_i]. \end{aligned}$$

Random assignment means $X_i \perp (\beta_{1i}, v_i)$, so $E[\beta_{1i} \mid X_i] = E[\beta_{1i}] = \beta_1$ and $E[v_i \mid X_i] = E[v_i]$. We assume without loss of generality that $E[v_i] = 0$ (any nonzero mean is absorbed into the intercept $\beta_0$). Therefore

$$E[u_i \mid X_i] = X_i \cdot (\beta_1 - \beta_1) + 0 = 0. \qquad \square$$

**c.** Since $E[u_i \mid X_i] = 0$ implies $\text{cov}(X_i, u_i) = 0$, the OLS slope converges to

$$\hat{\beta}_1 \xrightarrow{p} \frac{\text{cov}(X_i, Y_i)}{\text{Var}(X_i)} = \frac{\beta_1 \, \text{Var}(X_i) + \text{cov}(X_i, u_i)}{\text{Var}(X_i)} = \beta_1 = E[\beta_{1i}]. \qquad \square$$

**d.**

- *Assumption 1* ($E[u_i \mid X_i] = 0$): verified in (b). ✓
- *Assumption 2* ($(Y_i, X_i)$ i.i.d.): random sampling of individuals from the population and independent random assignment of treatment imply that observations are independent and identically distributed. ✓

Both assumptions are satisfied, so standard OLS inference applies.

**e.** Write:

$$E[u_i \mid X_i] = X_i \cdot (E[\beta_{1i} \mid X_i] - \beta_1) + E[v_i \mid X_i].$$

When $\text{cov}(\beta_{1i}, X_i) > 0$, individuals with high $X_i$ tend to have high $\beta_{1i}$, so $E[\beta_{1i} \mid X_i]$ is increasing in $X_i$. Hence $E[u_i \mid X_i] \neq 0$: **Assumption 1 is violated**.

Since $X_i$ is binary, the OLS slope with an intercept equals the difference in conditional means:

$$\hat{\beta}_1 \xrightarrow{p} E[Y_i \mid X_i = 1] - E[Y_i \mid X_i = 0].$$

Using $Y_i = \beta_0 + \beta_{1i} X_i + v_i$ and $E[v_i \mid X_i] = 0$,

$$\hat{\beta}_1 \xrightarrow{p} E[\beta_{1i} \mid X_i = 1].$$

Therefore the bias is

$$\hat{\beta}_1 \xrightarrow{p} E[\beta_{1i} \mid X_i = 1] \neq E[\beta_{1i}],$$

and more precisely,

$$\text{plim}\,\hat{\beta}_1 - E[\beta_{1i}] = E[\beta_{1i} \mid X_i = 1] - E[\beta_{1i}] = \frac{\text{cov}(\beta_{1i},\, X_i)}{P(X_i = 1)} > 0.$$

OLS is **upward-biased**: because high-gain individuals are disproportionately treated, the regression attributes their higher outcomes to the treatment indicator rather than to their underlying tendency to do better.

**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.

**Solution:**

**a.** Under monotonicity the three types are:

- **Always-takers** ($X_i = 1$ regardless of $Z_i$): identified by the treatment rate absent the lottery, $P(X_i = 1 \mid Z_i = 0)$.
- **Compliers** ($X_i = Z_i$): individuals whose treatment status is switched by the lottery.
- **Never-takers** ($X_i = 0$ regardless of $Z_i$): identified by the non-take-up rate even when selected, $P(X_i = 0 \mid Z_i = 1)$.

Proportions:

$$P(\text{AT}) = P(X_i = 1 \mid Z_i = 0) = 0.20,$$

$$P(\text{C}) = P(X_i = 1 \mid Z_i = 1) - P(X_i = 1 \mid Z_i = 0) = 0.60 - 0.20 = 0.40,$$

$$P(\text{NT}) = 1 - P(X_i = 1 \mid Z_i = 1) = 1 - 0.60 = 0.40.$$

Check: $0.20 + 0.40 + 0.40 = 1.00$. ✓

**b.** With a single binary instrument, TSLS reduces to the Wald estimator:

$$\hat{\beta}_1^{\text{TSLS}} = \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{P(X_i = 1 \mid Z_i = 1) - P(X_i = 1 \mid Z_i = 0)} = \frac{51{,}000 - 48{,}200}{0.6 - 0.2} = \frac{2{,}800}{0.4} = \boxed{7{,}000}.$$

**c.** By Equation (13.12), TSLS converges to

$$\hat{\beta}_1^{\text{TSLS}} \xrightarrow{p} \frac{E[\beta_{1i}\,\pi_{1i}]}{E[\pi_{1i}]},$$

a weighted average of individual effects, with weight $\pi_{1i}$ measuring how much the instrument shifts individual $i$'s treatment probability. For always-takers and never-takers, $\pi_{1i} = 0$ (the lottery does not change their behavior). For compliers, $\pi_{1i} > 0$. Therefore all weight falls on compliers and TSLS estimates the **local average treatment effect (LATE)**:

$$\text{LATE} = E[\beta_{1i} \mid \text{complier}] = 7{,}000.$$

The 40% of workers who enroll if and only if they win the lottery drive the entire estimate.

**d.** The population ATE weights all three types by their proportions:

$$
\begin{aligned}
\text{ATE} &= 0.20 \times 3{,}000 + 0.40 \times 7{,}000 + 0.40 \times 1{,}000 \\
&= 600 + 2{,}800 + 400 \\
&= \boxed{3{,}800}.
\end{aligned}
$$

To verify the Wald formula recovers the complier effect, note that under random assignment of $Z_i$:

$$
\begin{aligned}
E[Y_i \mid Z_i = 1] &= 0.20 \times E[Y_i(1) \mid \text{AT}] + 0.40 \times E[Y_i(1) \mid \text{C}] + 0.40 \times E[Y_i(0) \mid \text{NT}], \\
E[Y_i \mid Z_i = 0] &= 0.20 \times E[Y_i(1) \mid \text{AT}] + 0.40 \times E[Y_i(0) \mid \text{C}] + 0.40 \times E[Y_i(0) \mid \text{NT}].
\end{aligned}
$$

Subtracting: $E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] = 0.40 \times \left(E[Y_i(1) \mid \text{C}] - E[Y_i(0) \mid \text{C}]\right) = 0.40 \times 7{,}000 = 2{,}800$. Dividing by the first-stage 0.40 recovers LATE $= 7{,}000$. ✓

Since LATE $= 7{,}000 \neq 3{,}800 =$ ATE, the two parameters differ substantially.

**e.** A valid instrument shifts treatment for compliers but is invisible to always-takers and never-takers – their behavior does not change with $Z_i$, so they contribute nothing to the IV covariance. The instrument is effectively a random assignment device *only for compliers*. TSLS therefore learns exclusively about the complier subgroup.

The gap between LATE and ATE arises from **selection on gains**: workers self-select into enrollment partly based on how much they expect to benefit. Compliers – those who enroll when offered the chance – are precisely the workers who see training as sufficiently valuable, so they have above-average treatment effects (7,000 versus ATE = 3,800). TSLS cannot be "repaired" by choosing a better instrument: any valid binary instrument can only identify a LATE for its own complier group. In the language of Equation (13.12), because $\text{cov}(\beta_{1i}, \pi_{1i}) > 0$ – high-benefit workers are also the ones who respond most to incentives – the LATE $\neq$ ATE condition from Section 13.6 is violated.

**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 $\mathrm{Var}(Y_i(1))$ and $\mathrm{Var}(Y_i(0))$ are consistently estimable. Propose consistent estimators.

**d.** Show that

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

Hence argue that $\mathrm{Var}(\mathrm{TE}_i)$ 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:

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

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

**Solution:**

**a.** By random assignment, $X_i \perp (Y_i(0), Y_i(1))$, so

$$E[Y_i(1)] = E[Y_i(1) \mid X_i = 1] = E[Y_i \mid X_i = 1],$$

since $Y_i = Y_i(1)$ for treated units. By the law of large numbers, $\bar{Y}_{\text{treat}} = n_1^{-1} \sum_{X_i=1} Y_i \xrightarrow{p} E[Y_i \mid X_i = 1] = E[Y_i(1)]$. The same argument gives $\bar{Y}_{\text{control}} \xrightarrow{p} E[Y_i(0)]$. □

**b.** By linearity of expectations,

$$E[\mathrm{TE}_i] = E[Y_i(1)] - E[Y_i(0)] = E[Y_i \mid X_i = 1] - E[Y_i \mid X_i = 0].$$

By (a) and the continuous mapping theorem,

$$\bar{Y}_{\text{treat}} - \bar{Y}_{\text{control}} \xrightarrow{p} E[Y_i(1)] - E[Y_i(0)] = E[\mathrm{TE}_i]. \qquad \square$$

**c.** For any $k$, random assignment gives $E[Y_i(1)^k] = E[Y_i^k \mid X_i = 1]$. Therefore

$$\mathrm{Var}(Y_i(1)) = E[Y_i(1)^2] - (E[Y_i(1)])^2 = E[Y_i^2 \mid X_i = 1] - (E[Y_i \mid X_i = 1])^2 = \mathrm{Var}(Y_i \mid X_i = 1).$$

A consistent estimator is the within-treatment-arm sample variance:

$$\hat{\sigma}^2_{Y(1)} = \frac{1}{n_1 - 1} \sum_{X_i=1} (Y_i - \bar{Y}_{\text{treat}})^2 \xrightarrow{p} \text{Var}(Y_i(1)).$$

The same argument gives $\hat{\sigma}^2_{Y(0)} = (n_0 - 1)^{-1} \sum_{X_i=0}(Y_i - \bar{Y}_{\text{control}})^2 \xrightarrow{p} \text{Var}(Y_i(0))$. □

**d.** By the standard variance-of-a-difference identity:

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

Parts (a)–(c) show that $\sigma^2_{Y(1)}$ and $\sigma^2_{Y(0)}$ are identified and consistently estimable. The unidentified object is $\text{Cov}(Y_i(1), Y_i(0))$ – the covariance of the two potential outcomes **for the same individual**. From any experiment we observe at most one of $\{Y_i(0), Y_i(1)\}$ per unit. No matter how large the sample, the joint distribution of $(Y_i(0), Y_i(1))$ is never observed, and two joint distributions with identical marginals but different covariances are empirically indistinguishable. Therefore $\text{Cov}(Y_i(1), Y_i(0))$ – and hence $\text{Var}(\text{TE}_i)$ – cannot be consistently estimated. □

**e.** By the Cauchy-Schwarz inequality,

$$\left|\text{Cov}(Y_i(1),\, Y_i(0))\right| \le \sigma_{Y(1)}\,\sigma_{Y(0)},$$

so $\text{Cov}(Y_i(1), Y_i(0))$ lies in $[-\sigma_{Y(1)}\sigma_{Y(0)},\ \sigma_{Y(1)}\sigma_{Y(0)}]$.

*Upper bound on* $\text{Var}(\text{TE}_i)$: substitute the lower bound on the covariance,

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

*Lower bound on* $\text{Var}(\text{TE}_i)$: substitute the upper bound on the covariance,

$$\text{Var}(\text{TE}_i) \ge \sigma^2_{Y(1)} + \sigma^2_{Y(0)} - 2\sigma_{Y(1)}\sigma_{Y(0)} = \left(\sigma_{Y(1)} - \sigma_{Y(0)}\right)^2.$$

Both bounds are functions of $\sigma_{Y(1)}$ and $\sigma_{Y(0)}$ alone, which are consistently estimable by (c), so the bounds themselves are consistently estimable. □

*Interpretation.* These are sharp **variance-based bounds**: given only $\sigma_{Y(1)}$ and $\sigma_{Y(0)}$, Cauchy-Schwarz cannot improve them. The **lower bound** $(\sigma_{Y(1)} - \sigma_{Y(0)})^2$ is attained when the standardized potential outcomes are perfectly positively correlated, and the **upper bound** $(\sigma_{Y(1)} + \sigma_{Y(0)})^2$ when they are perfectly negatively correlated – whenever such joint distributions are feasible given the two marginals. In the extreme case $\sigma_{Y(1)} = \sigma_{Y(0)}$, the lower bound is zero, meaning the data are consistent with a constant treatment effect for all individuals, though they do not require it. The gap between the two bounds is $4\sigma_{Y(1)}\sigma_{Y(0)}$, which is large whenever both arms are individually variable: the experiment pins down the mean treatment effect precisely but leaves substantial uncertainty about how heterogeneous it is.