Microeconometrics II — Problem Set 01 — LATE

Vinícius de Almeida Nery Ferreira

October 17th, 2024

1 One-Sided Non-Compliance

We saw in class that the Wald estimand

$$\beta_W = \frac{\mathbb{E}[Y \mid Z=1] - \mathbb{E}[Y \mid Z=0]}{\mathbb{E}[D \mid Z=1] - \mathbb{E}[D \mid Z=0]} \tag{Wald}$$

identifies the average treatment effect for compliers (LATE)

$$LATE = \mathbb{E}[Y(1) - Y(0) \mid D(0)=0, D(1)=1] \tag{LATE}$$

under the exclusion, exogeneity and monotonicity assumptions and when D and Z are binary variables and we have no additional covariates.

Suppose now that we have one-sided non-compliance. This means that we don't have always-takers: individuals only receive treatment if they were incentivized to do so, but are not mandated to take treatment if they are assigned the instrument. Mathematically, this means $\mathbb{P}(D(0)=1)=0$.

The average treatment effect on the treated (ATT) is defined as

$$ATT := \mathbb{E}[Y(1) - Y(0) \mid D=1]$$

Note that, by the LIE, we can write

$$\begin{aligned}\mathbb{E}[Y(1) - Y(0) \mid D=1] = &\mathbb{P}(Z=0 \mid D=1) \cdot \mathbb{E}[Y(1) - Y(0) \mid D=1, Z=0] \\ &+ \mathbb{P}(Z=1 \mid D=1) \cdot \mathbb{E}[Y(1) - Y(0) \mid D=1, Z=1]\end{aligned}$$

By one-sided non-compliance, we know that $\mathbb{P}(Z=0 \mid D=1)=0$: only those that were incentivized can receive treatment. By the Law of Total Probability, this implies $\mathbb{P}(Z=1 \mid D=1)=1$. Using this fact and substituting potential outcomes, we get

$$\mathbb{E}[Y(1) - Y(0) \mid D=1] = \mathbb{E}[Y(1) - Y(0) \mid D(1)=1, Z=1]$$

Finally, note that the events $\{D(1)=1, Z=1\}$ and $\{D(1)=1, D(0)=0\}$ are equivalent in our context[1], since we don't have always-takers: if someone is treated, they must be a complier. Thus,

$$\mathbb{E}[Y(1) - Y(0) \mid D=1] = \mathbb{E}[Y(1) - Y(0) \mid D(0)=0, D(1)=1] \quad \therefore \quad \boldsymbol{ATT = LATE = \beta_W},$$

where the last equality comes from the identification result we saw in class.

[1] Formally, we would have to show set equivalence, but I think we are past this point.

2 IV Regression with Covariates

We will prove Theorem 3.3. and Corollary 3.4. of Słoczyński (2024). We follow the proof in the appendix of an earlier version of the paper.

For that, the following lemma (stated without proof in the paper) will be useful.

Lemma. *Consider the model*

$$Y = \alpha + \tau D + X'\gamma + \varepsilon,$$

where α, τ and γ are OLS coefficients, Y is the (scalar) dependent variable, D is a binary variable, X is a $K \times 1$ vector of covariates and ε is an error term. Define

$$\tau(x) = \mathbb{E}\left[Y \mid D = 1, X = x\right] - \mathbb{E}\left[Y \mid D = 0, X = x\right]$$

Assume that $\mathbb{E}[D \mid X]$ is linear in X. Then, the OLS coefficient τ can be written as

$$\tau = \frac{\mathbb{E}_X\left[\mathrm{Var}(D \mid X) \cdot \tau(x)\right]}{\mathbb{E}_X\left[\mathrm{Var}(D \mid X)\right]}$$

Proof. Note that, by the Frisch-Waugh-Lovell Theorem,

$$\tau = \frac{\mathrm{Cov}(\tilde{D}, Y)}{\mathrm{Var}(\tilde{D})}$$

where $\tilde{D} \coloneqq D - \mathbb{E}[D \mid X]$ is the residual of a regression of D on X. Note that, by the assumption of linearity of $\mathbb{E}[D \mid X]$, $\mathbb{E}[\tilde{D}] = \mathbb{E}[\tilde{D} \mid X] = 0$.

Note now that we can decompose Y as

$$\begin{aligned}
Y &= \mathbb{E}[Y \mid D, X = x] + \varepsilon \\
&= \mathbb{E}[Y \mid D = 0, X = x] + (\mathbb{E}[Y \mid D = 1, X = x] - \mathbb{E}[Y \mid D = 0, X = x]) \cdot D + \varepsilon \\
&= \mathbb{E}[Y \mid D = 0, X = x] + \tau(x) \cdot D + \varepsilon
\end{aligned}$$

Substituting this expression for Y in the expression for τ,

$$\begin{aligned}
\tau &= \frac{\mathrm{Cov}(\tilde{D}, Y)}{\mathrm{Var}(\tilde{D})} \\
&= \frac{\mathrm{Cov}\left(\tilde{D}, \mathbb{E}[Y \mid D = 0, X = x] + \tau(x) \cdot D + \varepsilon\right)}{\mathrm{Var}\left(\tilde{D}\right)} \\
&= \frac{\mathrm{Cov}\left(\tilde{D}, \mathbb{E}[Y \mid D = 0, X = x]\right)}{\mathrm{Var}\left(\tilde{D}\right)} + \frac{\mathrm{Cov}\left(\tilde{D}, \tau(x) \cdot D\right)}{\mathrm{Var}\left(\tilde{D}\right)} + \frac{\mathrm{Cov}\left(\tilde{D}, \varepsilon\right)}{\mathrm{Var}\left(\tilde{D}\right)}
\end{aligned}$$

Since $\tilde{D}$ and ε are OLS errors which are both (partial) functions of D and X, the third term is zero. Note that this is also the case for the first term, since, by the LIE and linearity ($\mathbb{E}\left[\tilde{D} \mid X\right] = 0$),

$$\begin{aligned}
\operatorname{Cov}\left(\tilde{D}, \mathbb{E}[Y \mid D=0, X=x]\right) &= \mathbb{E}\left[\tilde{D} \cdot \mathbb{E}[Y \mid D=0, X=x]\right] \\
&= \mathbb{E}\left\{\mathbb{E}\left[\tilde{D} \cdot \mathbb{E}[Y \mid D=0, X=x] \mid X\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}[Y \mid D=0, X=x] \cdot \mathbb{E}\left[\tilde{D} \mid X\right]\right\} \\
&= 0
\end{aligned}$$

For the second term, note that

$$\begin{aligned}
\frac{\operatorname{Cov}\left(\tilde{D}, \tau(x) \cdot D\right)}{\operatorname{Var}\left(\tilde{D}\right)} &= \frac{\operatorname{Cov}\left(\tilde{D}, \tau(x) \cdot \left(\tilde{D} + \mathbb{E}[D \mid X]\right)\right)}{\operatorname{Var}\left(\tilde{D}\right)} \\
&= \frac{\mathbb{E}\left[\tilde{D}^2 \tau(x)\right] + \mathbb{E}\left[\tilde{D} \cdot \tau(x) \cdot E[D \mid X]\right]}{\mathbb{E}\left[\tilde{D}^2\right]}
\end{aligned}$$

Note that the second part of the enumerator of this expression is zero:

$$\begin{aligned}
\mathbb{E}\left[\tilde{D} \cdot \tau(x) \cdot E[D \mid X]\right] &= \mathbb{E}\left[(D - E[D \mid X]) \cdot \tau(x) \cdot E[D \mid X]\right] \\
&= \mathbb{E}_X\left\{\mathbb{E}\left[(D - E[D \mid X]) \cdot \tau(x) \cdot E[D \mid X] \mid X\right]\right\} \\
&= \mathbb{E}_X\left\{\tau(x) \cdot E[D \mid X] \cdot \mathbb{E}\left[(D - E[D \mid X]) \mid X\right]\right\} \\
&= \mathbb{E}_X\left\{\tau(x) \cdot E[D \mid X] \cdot \mathbb{E}\left[\tilde{D} \mid X\right]\right\} \\
&= 0
\end{aligned}$$

since $\mathbb{E}\left[\tilde{D} \mid X\right] = 0$ by the linearity assumption. We then have the desired result:

$$\begin{aligned}
\tau = \frac{\operatorname{Cov}(\tilde{D}, Y)}{\operatorname{Var}(\tilde{D})} &= \frac{\mathbb{E}\left[\tilde{D}^2 \tau(x)\right]}{\mathbb{E}\left[\tilde{D}^2\right]} \\
&= \frac{\mathbb{E}_X\left\{\mathbb{E}\left[\tilde{D}^2 \tau(x) \mid X\right]\right\}}{\mathbb{E}_X\left\{\mathbb{E}\left[\tilde{D}^2 \mid X\right]\right\}} \\
&= \frac{\mathbb{E}_X\left\{\tau(x) \cdot \mathbb{E}\left[(D - \mathbb{E}[D \mid X])^2 \mid X\right]\right\}}{\mathbb{E}_X\left\{\mathbb{E}\left[(D - \mathbb{E}[D \mid X])^2 \mid X\right]\right\}} \\
&= \frac{\mathbb{E}_X\left\{\tau(x) \cdot \operatorname{Var}(D \mid X)\right\}}{\mathbb{E}_X\left\{\operatorname{Var}(D \mid X)\right\}}
\end{aligned}$$

□

2.1 Proof of Theorem 3.3.

Theorem 3.3. Suppose

1. **Assumption IV**
   i. *(Conditional Independence)*: $(Y(0,0), Y(0,1), Y(1,0), Y(1,1), D(0), D(1)) \perp Z \mid X$;
   ii. *(Exclusion Restriction)*: $\mathbb{P}[Y(1,d) = Y(0,d) \mid X] = 1$ for $d \in \{0,1\}$ a.s.;
   iii. *(Relevance)*: $0 < \mathbb{P}[Z = 1 \mid X] < 1$ and $\mathbb{P}[D(1) = 1 \mid X] \neq \mathbb{P}[D(0) = 1 \mid X]$ a.s.;
2. **Assumption WM** *(Weak Monotonicity)*: There exists a subset of the support of X such that $\mathbb{P}[D(1) \geq D(0) \mid X] = 1$ on it and $\mathbb{P}[D(1) \leq D(0) \mid X] = 1$ on its complement;
3. **Assumption PS** *(Instrument Propensity Score)*: $e(X) = \mathbb{E}[Z \mid X] = X\alpha$.

   Then,

$$\beta_{\mathrm{IV}} = \frac{\mathbb{E}[c(X) \cdot \pi(X) \cdot \mathrm{Var}[Z \mid X] \cdot \tau(X)]}{\mathbb{E}[c(X) \cdot \pi(X) \cdot \mathrm{Var}[Z \mid X]]},$$

where

$$\beta_{\mathrm{IV}} = \left[\mathbb{E}\left[Q'W\right]^{-1} \mathbb{E}\left[Q'Y\right]\right]_1$$

is the usual IV estimand[2], with $W = (D, X)$ and $Q = (Z, X)$, where D is the binary treatment variable, Z is the binary instrument and X is a $1 \times K$ row vector of covariates, which includes 1;

$$c(x) = \mathrm{sgn}\left(\mathbb{P}\left(D(1) \geq D(0) \mid X = x\right) - \mathbb{P}\left(D(1) \leq D(0) \mid X = x\right)\right)$$

with $\mathrm{sgn}(\cdot) \in \{-1, 1\}$ being the sign function;

$$\tau(x) = \mathbb{E}\left[Y(1) - Y(0) \mid D(1) \neq D(0), X = x\right]$$

is the LATE conditional on $X = x$; and

$$\pi(x) = \mathbb{P}\left[D(1) \neq D(0) \mid X = x\right]$$

is the conditional proportion of compliers or defiers.

Proof. Recall that β_{IV} is equal to the ratio of the reduced-form and first-stage coefficients OLS of Z:

$$Y = \phi Z + X\gamma_2 + \eta_2$$

$$D = \omega Z + X\gamma_1 + \eta_1$$

$$\beta_{\mathrm{IV}} = \frac{\phi}{\omega}$$

Using our lemma separately in each of these coefficients – which is possible by Assumption PS, which states that $\mathbb{E}[Z \mid X]$ is linear in X –, we can write

$$\beta_{\mathrm{IV}} = \frac{\frac{\mathbb{E}_X\{\phi(x) \cdot \mathrm{Var}(Z|X=x)\}}{\mathbb{E}_X\{\mathrm{Var}(Z|X=x)\}}}{\frac{\mathbb{E}_X\{\omega(x) \cdot \mathrm{Var}(Z|X=x)\}}{\mathbb{E}_X\{\mathrm{Var}(Z|X=x)\}}} \tag{1}$$

[2]Note that we are only interested in the first element of the vector $\mathbb{E}\left[Q'W\right]^{-1} \mathbb{E}\left[Q'Y\right]$. In other words, we are interested in the coefficient β of $Y = \beta D + X\gamma + \eta$.

where

$$\phi(x) = \mathbb{E}[Y \mid Z=1, X=x] - \mathbb{E}[Y \mid Z=0, X=x]$$
$$\omega(x) = \mathbb{E}[D \mid Z=1, X=x] - \mathbb{E}[D \mid Z=0, X=x]$$

are the conditional slope coefficients of the respective regression. Note that $\beta_{\text{IV}} = \frac{\phi}{\omega}$ is also valid on a conditional basis, so that

$$\beta(x) = \frac{\phi(x)}{\omega(x)} \quad \therefore \quad \phi(x) = \beta(x) \cdot \omega(x)$$

We can rearrange Equation 1 to get

$$\begin{aligned}\beta_{\text{IV}} &= \frac{\mathbb{E}_X\left\{\phi(x) \cdot \text{Var}(Z \mid X=x)\right\}}{\mathbb{E}_X\left\{\omega(x) \cdot \text{Var}(Z \mid X=x)\right\}} \\ &= \frac{\mathbb{E}_X\left\{\beta(x) \cdot \omega(x) \cdot \text{Var}(Z \mid X=x)\right\}}{\mathbb{E}_X\left\{\omega(x) \cdot \text{Var}(Z \mid X=x)\right\}}\end{aligned}$$

We know that, under Assumptions IV and WM, the IV estimand identifies the LATE for those whose treatment status is impacted by the instrument. A similar argument shows that

$$\beta(x) = \tau(x)$$

Under those same assumptions, we can identify the proportion of the population whose treatment status changes with the instrument:

$$\pi(x) = \begin{cases} \omega(x), & \mathbb{P}[D(1) \geq D(0) \mid X] = 1 \text{ (Compliers)} \\ -\omega(x), & \mathbb{P}[D(1) \leq D(0) \mid X] = 1 \text{ (Defiers)} \end{cases}$$

We can write this more elegantly as $\omega(x) = c(x) \cdot \pi(x)$. Substituting these expression into β_{IV},

$$\begin{aligned}\beta_{\text{IV}} &= \frac{\mathbb{E}_X\left\{\beta(x) \cdot \omega(x) \cdot \text{Var}(Z \mid X=x)\right\}}{\mathbb{E}_X\left\{\omega(x) \cdot \text{Var}(Z \mid X=x)\right\}} \\ &= \frac{\mathbb{E}_X\left\{\tau(x) \cdot c(x) \cdot \pi(x) \cdot \text{Var}(Z \mid X=x)\right\}}{\mathbb{E}_X\left\{c(x) \cdot \pi(x) \cdot \text{Var}(Z \mid X=x)\right\}} \\ &= \frac{\mathbb{E}_X\left\{c(x) \cdot \pi(x) \cdot \text{Var}(Z \mid X=x) \cdot \tau(x)\right\}}{\mathbb{E}_X\left\{c(x) \cdot \pi(x) \cdot \text{Var}(Z \mid X=x)\right\}}\end{aligned}$$

□

As discussed in Słoczyński (2024), the usual IV application does not include interaction terms of the instrument with the covariates in neither stage, which restricts the effect of Z to be homogeneous across covariate values. If that is the case, then we don't necessarily have that β_{IV} is a *convex* combination of conditional LATEs $\tau(x)$.

This is because $c(x)$ is negative for every covariate value where there exist defiers (but no compliers). Thus, in the usual application of IV and under the assumptions of the theorem – crucially, under WM –, the IV estimand may no longer have a causal interpretation: it is even possible that β_{IV} be negative (positive) when treatment effects are positive (negative) for everyone in the population!

2.2 Proof of Corollary 3.4.

Corollary 3.4. Suppose

1. **Assumption IV**
   i. *(Conditional Independence)*: $(Y(0,0), Y(0,1), Y(1,0), Y(1,1), D(0), D(1)) \perp Z \mid X$;
   ii. *(Exclusion Restriction)*: $\mathbb{P}[Y(1,d) = Y(0,d) \mid X] = 1$ for $d \in \{0,1\}$ a.s.;
   iii. *(Relevance)*: $0 < \mathbb{P}[Z = 1 \mid X] < 1$ and $\mathbb{P}[D(1) = 1 \mid X] \neq \mathbb{P}[D(0) = 1 \mid X]$ a.s.;
2. **Assumption SM** *(Strong Monotonicity)*: $\mathbb{P}[D(1) \geq D(0) \mid X] = 1$ almost surely;
3. **Assumption PS** *(Instrument Propensity Score)*: $e(X) = \mathbb{E}[Z \mid X] = X\alpha$.

Then,

$$\beta_{\text{IV}} = \frac{\mathbb{E}[\pi(X) \cdot \text{Var}[Z \mid X] \cdot \tau(X)]}{\mathbb{E}[\pi(X) \cdot \text{Var}[Z \mid X]]},$$

Proof. All steps are similar to those of the proof of Theorem 3.3. The distinction is that, as we are under Assumption SM, we rule out defiers at every covariate value, so that

$$\pi(x) = \omega(x)$$

We then have that

$$\begin{aligned}
\beta_{\text{IV}} &= \frac{\mathbb{E}_X \{\beta(x) \cdot \omega(x) \cdot \text{Var}(Z \mid X = x)\}}{\mathbb{E}_X \{\omega(x) \cdot \text{Var}(Z \mid X = x)\}} \\
&= \frac{\mathbb{E}_X \{\tau(x) \cdot \pi(x) \cdot \text{Var}(Z \mid X = x)\}}{\mathbb{E}_X \{\pi(x) \cdot \text{Var}(Z \mid X = x)\}} \\
&= \frac{\mathbb{E}_X \{\pi(x) \cdot \text{Var}(Z \mid X = x) \cdot \tau(x)\}}{\mathbb{E}_X \{\pi(x) \cdot \text{Var}(Z \mid X = x)\}}
\end{aligned}$$

□

With strong monotonicity, the IV estimand retains the property of being a convex combination of conditional LATEs. Note that, the higher the variance of Z for $X = x$ or the larger the proportion of compliers, the bigger the weight for the group $X = x$.