PS690 Introduction to Regression Methods Spring 2026

Lecture 7 — Instrumental Variables

Jiawei Fu, Duke University *Scribe: Jiawei Fu*

1 Overview

In previous lectures, when studying structural models, we typically assumed that there was no endogeneity problem. That assumption is what allowed OLS to identify the parameters of interest. In many empirical settings, however, this assumption is not plausible. For example, researchers often cannot measure all relevant omitted variables, so those variables cannot be controlled for directly. In that case, the OLS estimator is generally inconsistent. In this lecture, we study an alternative identification strategy based on instrumental variables.

2 Smearing

Let us first see what happens to the OLS estimator if one regressor is endogenous.

Previously, we assumed $\mathbb{E}[X_i e_i] = 0$. Now suppose instead that $\mathbb{E}[X_i e_i] = \Delta$, where Δ is a nonzero vector because X_i and e_i are correlated.

Consider $Y = X\beta + e$.

$$\begin{aligned}\hat{\beta}_{ols} &= \beta + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i e_i) \\ &\to \beta + (\mathbb{E}[X_i X_i'])^{-1}\Delta\end{aligned}$$

Now suppose that only the last variable, X_K, is endogenous. Then

$$\Delta = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \delta \end{bmatrix}.$$

It follows that

$$\hat{\beta}_{ols} \to \beta + \delta \times \text{the last column of } (\mathbb{E}[X_i X_i'])^{-1}$$

There is no reason to expect any element of the last column of $(\mathbb{E}[X_i X_i'])^{-1}$ to equal zero. The implication is that even though only one regressor in X is correlated with e, all elements of $\hat{\beta}$ are generally inconsistent, not just the coefficient on the endogenous regressor. This effect is called *smearing*: the inconsistency caused by one endogenous variable is smeared across all least-squares coefficients.

3 Identification and Estimation

We will develop the IV framework in three steps, moving from the simplest case to the general case.

3.1 Single Endogenous Variable

Consider

$$Y_i = \beta_0 + \beta_1 X_{1i} + e_i,$$

where $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[X_{1i} e_i] \neq 0$. In this case, X_{1i} is endogenous, so β_1 is not identified by the usual OLS moment condition.

Now suppose we have another variable Z_i such that $\mathbb{E}[Z_i e_i] = 0$. We will use Z_i as an instrument. Multiplying both sides by Z_i and taking covariance yields

$$\begin{aligned} Cov(Z_i, Y_i) &= Cov(Z_i, \beta_0) + \beta_1 Cov(Z_i, X_{1i}) + Cov(Z_i, e_i) \\ Cov(Z_i, Y_i) &= \beta_1 Cov(Z_i, X_{1i}) \\ \beta_1 &= \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_{1i})} \end{aligned}$$

provided that $Cov(Z_i, X_{1i}) \neq 0$.

This simple example illustrates the two key conditions for a valid instrument:

1. **Exogeneity (or exclusion):** $Cov(Z_i, e_i) = 0$.

2. **Relevance:** $Cov(Z_i, X_{1i}) \neq 0$.

Standardize the numerator and denominator by $Var(Z_i)$ to obtain

$$\beta_1 = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, X_{1i})/Var(Z_i)}$$

The numerator is the BLP coefficient from regressing Y_i on Z_i, and the denominator is the BLP coefficient from regressing X_{1i} on Z_i. In other words, the IV estimand is the ratio of the reduced-form effect of Z_i on Y_i to the first-stage effect of Z_i on X_{1i}.

If Z_i is binary, these coefficients are just differences in means, so

$$\beta_1 = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[X_{1i} | Z_i = 1] - \mathbb{E}[X_{1i} | Z_i = 0]}$$

This is called the Wald estimator. Each conditional mean can be estimated by a sample average such as

$$\frac{\sum_{i=1}^n Z_i W_i}{\sum_{i=1}^n Z_i}$$

3.2 Single Endogenous Variable and Exogenous Variables

Consider

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1,i} + e_i = X_i' \beta + e_i,$$

where $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[X_{1i}e_i] \neq 0$.

In this case, we need an instrument Z_{1i} such that $Cov(Z_{1i}, e_i) = \mathbb{E}[Z_{1i}e_i] = 0$. This is the exclusion restriction. We also need relevance, but here relevance must be understood conditional on the other exogenous variables. A precise statement comes from the reduced form for X_{1i}:

$$X_{1i} = \delta_0 + \delta_1 Z_{1i} + \delta_2 X_{2i} + ... + \delta_{k-1} X_{k-1,i} + u_i.$$

The key condition is $\delta_1 \neq 0$. This is often described loosely as saying that Z_{1i} is correlated with X_{1i}, but that wording is incomplete. More precisely, Z_{1i} must be partially correlated with X_{1i} after the other exogenous variables have been partialled out. Unlike exclusion, this relevance condition can be assessed empirically.

Now we can show that the whole parameter vector β is identified. Write

$$Z_i = (1, Z_{1i}, X_{2i}, ..., X_{k-1,i})^T.$$

Note that we include the exogenous regressors themselves in the instrument vector, since exogenous variables are valid instruments for themselves.

$$\begin{aligned}
0 &= \mathbb{E}[Z_i e_i] \qquad \text{(exclusion restriction)} \\
0 &= \mathbb{E}[Z_i(Y_i - X_i'\beta)] \\
0 &= \mathbb{E}[Z_i Y_i] - \mathbb{E}[Z_i X_i']\beta \\
\beta &= (\mathbb{E}[Z_i X_i'])^{-1}\mathbb{E}[Z_i Y_i]
\end{aligned}$$

For the last step, we need $\mathbb{E}[Z_i X_i']$ to be invertible; equivalently, we need the rank condition. In this just-identified case, that condition follows if $\delta_1 \neq 0$.

The IV estimator is

$$\begin{aligned}
\hat{\beta}_{iv} &= (\frac{1}{n}\sum_{i=1}^{n} Z_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i) \\
&= (Z'X)^{-1}(Z'Y)
\end{aligned}$$

3.3 Multiple Instruments: Two-Stage Least Squares

Let us continue to consider

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1,i} + e_i = X_i'\beta + e_i,$$

where $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[X_{1i}e_i] \neq 0$.

Now suppose we have more than one valid instrument for X_{1i}. Let $Z_{1i}, ..., Z_{Mi}$ satisfy $Cov(Z_{hi}, e_i) = 0$ for $h = 1, 2, ..., M$. If each instrument has some partial correlation with X_{1i}, then in principle we could form many different IV estimators. In fact, there are infinitely many, because any linear combination of $X_{2i}, ..., X_{k-1,i}, Z_{1i}, ..., Z_{Mi}$ is also uncorrelated with e_i. So which IV estimator should we use? This is the over-identified case, because we have more instruments than endogenous regressors. The previous case with one instrument was just identified.

Define

$$Z_i = (1, Z_{1i}, ..., Z_{Mi}, X_{2i}, ...X_{k-1,i})^T,$$

an $L \times 1$ vector of instruments. Consider the reduced form for X_{1i}:

$$\begin{aligned} X_{1i} &= \gamma_0 + \gamma_1 Z_{1i} + ... + \gamma_M Z_{Mi} + \gamma_{M+1} X_{2i} + ... + \gamma_{M+k-1} X_{k-1,i} + u_i \\ X_{1i} &= Z_i' \Gamma + u_i \end{aligned}$$

We know that $\Gamma = \mathbb{E}\left[Z_i Z_i'\right]^{-1} \mathbb{E}\left[Z_i X_{1i}\right]$. The population fitted value is

$$X_{1i}^* = Z_i' \mathbb{E}\left[Z_i Z_i'\right]^{-1} \mathbb{E}\left[Z_i X_{1i}\right].$$

Now substitute $X_{1i} = X_{1i}^* + u_i$ into the structural equation for Y_i:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1,i} + e_i \\ Y_i &= \beta_0 + \beta_1 (X_{1i}^* + u_i) + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1,i} + e_i \\ Y_i &= \beta_0 + \beta_1 (\Gamma' Z_i) + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1,i} + (e_i + \beta_1 u_i) \end{aligned}$$

Note that $\mathbb{E}\left[(\Gamma' Z_i)(e_i + \beta_1 u_i)\right] = \Gamma' \mathbb{E}\left[Z_i e_i\right] + \beta_1 \Gamma' \mathbb{E}\left[Z_i u_i\right] = 0$. Therefore, the fitted value X_{1i}^* is exogenous, which suggests estimating the model by replacing X_{1i} with its projection on the instrument space.

Now let us move to the sample analogue:

$$Y = \beta_0 + \beta_1 (Z\Gamma) + \beta_2 X_2 + ... + \beta_{k-1} X_{k-1} + (e + \beta_1 u)$$

Γ is not observed, but we can estimate it by OLS:

$$\hat{\Gamma} = (Z'Z)^{-1} Z' X_1.$$

The fitted value of X_1 is

$$\hat{X}_1 = Z\hat{\Gamma} = Z(Z'Z)^{-1} Z' X_1.$$

Define the $n \times k$ matrix $\hat{X} = (1, \hat{X}_1, X_2, ..., X_{k-1})$. Then

$$\begin{aligned} Y &= \beta_0 + \beta_1 (Z\hat{\Gamma}) + \beta_2 X_2 + ... + \beta_{k-1} X_{k-1} + v \\ Y &= \beta_0 + \beta_1 \hat{X}_1 + \beta_2 X_2 + ... + \beta_{k-1} X_{k-1} + v \\ Y &= \hat{X}\beta + v \end{aligned}$$

The OLS estimator from this transformed equation is

$$\hat{\beta}_{2sls} = (\hat{X}'\hat{X})^{-1} \hat{X}' Y.$$

It is called the two-stage least squares (2SLS) estimator. In summary, $\hat{\beta}_{2sls}$ is obtained by:

1. **First stage:** regress X_1 on $1, Z_1, ..., Z_M, X_2, ..., X_{k-1}$ and obtain the fitted values $\hat{X}_1$.

2. **Second stage:** regress Y on $1, \hat{X}_1, X_2, ..., X_{k-1}$.

You should include all exogenous variables in both stages. In practice, do not run the two stages "by hand" if your software has a built-in 2SLS routine, because the standard errors must account for the fact that the first stage is estimated.

The 2SLS estimator also has several equivalent representations. Note that

$$\hat{X} = Z(Z'Z)^{-1}Z'X = P_Z X,$$

where P_Z is the projection matrix onto the column space of Z. For exogenous regressors, projecting onto the instrument space leaves them unchanged. Therefore,

$$\begin{aligned}\hat{\beta}_{2sls} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y \\ &= (X'P_Z P_Z X)^{-1}\hat{X}'Y \\ &= (X'P_Z X)^{-1}\hat{X}'Y \\ &= (\hat{X}'X)^{-1}\hat{X}'Y\end{aligned}$$

This shows that 2SLS is an IV estimator that uses the fitted value $\hat{X}_1$ as the instrument for X_1. Since X_{1i}^* is a linear combination of exogenous variables, it is itself a valid instrument.

Under the just-identified case, $L = K$ and $\hat{\beta}_{2sls} = \hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$.

Also, you can check we can also get $\hat{\beta}_{2sls} = (X'P_Z X)^{-1}(X'P_Z Y)$.

4 Consistency and Inference

Assumption 1 (Exclusion Restriction)**.** *For some $L \times 1$ vector Z_i, $\mathbb{E}[Z_i e_i] = 0$*

Unless every element of X_i is exogenous, Z_i must contain at least some variables obtained from outside the structural equation.

Assumption 2. *(a)$rank(\mathbb{E}[ZZ']) = L$ (b)$rank(\mathbb{E}[ZX']) = k$ (relevance condition or rank condition)*

A necessary condition for $rank(\mathbb{E}[ZX']) = k$ is the order condition $L \geq k$. In other words, we must have at least as many instruments as endogenous-plus-exogenous regressors in the structural equation. We say that the model is **just-identified** if $L = k$ and **over-identified** if $L > k$. Let us now show that $\hat{\beta}_{2sls}$ is consistent.

Note $\hat{\beta}_{2sls} = (X'P_Z X)^{-1}X'P_Z Y$. Now, recall $Y = X\beta + e$. Therefore,

$$\begin{aligned}\hat{\beta}_{2sls} &= (X'P_Z X)^{-1}X'P_Z Y \\ &= (X'P_Z X)^{-1}X'P_Z(X\beta + e) \\ &= \beta + (X'P_Z X)^{-1}X'P_Z e \\ &= \beta + [(\frac{1}{n}X'Z)(\frac{1}{n}Z'Z)^{-1}(\frac{1}{n}Z'X)]^{-1}(\frac{1}{n}X'Z)(\frac{1}{n}Z'Z)^{-1}(\frac{1}{n}Z'e) \\ &\to \beta + (\mathbb{E}[X_i Z_i'](\mathbb{E}[Z_i Z_i'])^{-1}\mathbb{E}[Z_i X_i'])^{-1}\mathbb{E}[X_i Z_i'](\mathbb{E}[Z_i Z_i'])^{-1}\mathbb{E}[Z_i e_i] \\ &= \beta\end{aligned}$$

The matrix $(\mathbb{E}[X_i Z_i'](\mathbb{E}[Z_i Z_i'])^{-1}\mathbb{E}[Z_i X_i'])^{-1}$ exists under the rank condition.

Unlike OLS under strict exogeneity, 2SLS is generally not unbiased in finite samples. What we typically rely on is consistency and asymptotic normality.

Note that $\frac{1}{\sqrt{n}}Z'e \to N(0,\Omega)$, $\Omega = \mathbb{E}\left[Z_i Z_i e_i^2\right]$. Then, similar to OLS, we can show

$$\sqrt{n}(\hat{\beta}_{2sls} - \beta) \to N(0, V_\beta)$$

where

$$\begin{aligned} V_\beta =& (\mathbb{E}\left[X_i Z_i'\right](\mathbb{E}\left[Z_i Z_i'\right])^{-1}\mathbb{E}\left[Z_i X_i'\right])^{-1} \\ & \times (\mathbb{E}\left[X_i Z_i'\right](\mathbb{E}\left[Z_i Z_i'\right])^{-1}\Omega(\mathbb{E}\left[Z_i Z_i'\right])^{-1}\mathbb{E}\left[Z_i X_i'\right]) \\ & \times (\mathbb{E}\left[X_i Z_i'\right](\mathbb{E}\left[Z_i Z_i'\right])^{-1}\mathbb{E}\left[Z_i X_i'\right])^{-1}, \end{aligned}$$

and $\Omega = \mathbb{E}\left[Z_i Z_i' e_i^2\right]$.

Under homoskedasticity, $\Omega = \mathbb{E}\left[Z_i Z_i'\right]\sigma^2$, so

$$V_\beta = (\mathbb{E}\left[X_i Z_i'\right](\mathbb{E}\left[Z_i Z_i'\right])^{-1}\mathbb{E}\left[Z_i X_i'\right])^{-1}\sigma^2$$

Thus the asymptotic variance is larger when the structural disturbance is more variable and smaller when the instruments are more informative about the endogenous regressors.

Theorem 3. *Under Assumptions 1,2 and homoskedasticity, the 2SLS estimator is efficient in the class of all instrumental variables estimators using instruments linear in Z_i.*

Asymptotically, under homoskedasticity, it is beneficial to use as many valid instruments as are available. Using only a subset of Z amounts to using a particular linear combination of the full instrument set. For some subsets we may attain the same efficiency as 2SLS with all instruments, but we cannot do better. This observation makes it tempting to include many instruments so that L is much larger than K. Unfortunately, there is a catch: the finite-sample bias of 2SLS generally increases with the number of instruments. This is one aspect of the many-instruments problem.

How do we estimate Ω? A natural estimator uses the 2SLS residuals:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{e}_i^2.$$

But be careful: the residual here is

$$\hat{e}_i = Y_i - X_i'\hat{\beta}_{2sls}$$

not the residual from the hand-run second-stage regression $Y_i - \hat{X}_i'\hat{\beta}$.

5 Weak Identification

In this section, we follow Keane & Neal (2024).

5.1 Identification Failure

Consider the simplest case with one endogenous regressor and one instrument:

$$Y = X\beta + e$$
$$X = Z\gamma + u$$

where $\gamma = \frac{\mathbb{E}[ZX]}{\mathbb{E}[Z^2]}$. Suppose $\gamma = 0$, that is, $\mathbb{E}[ZX] = 0$. Then the instrument has no relevance, and the IV estimator is

$$\hat{\beta}_{iv} - \beta = \frac{\sum_{i=1}^n Z_i e_i}{\sum_{i=1}^n Z_i X_i} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i e_i}{\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i u_i} \to \frac{\xi_e}{\xi_u}$$

where, by the CLT, $(\xi_e, \xi_u) \sim N(0, \begin{bmatrix} \sigma_e^2 & \sigma_{eu} \\ \sigma_{eu} & \sigma_u^2 \end{bmatrix}$. Let $\delta = \frac{\sigma_{eu}}{\sigma_u^2}$, then $\xi = \xi_e - \delta\xi_u$ is normal and independent of ξ_u. Then,

$$\hat{\beta}_{iv} - \beta \to \delta + \frac{\xi}{\xi_u}$$

So $\hat{\beta}_{iv}$ does not converge in probability to a constant; instead, it converges in distribution to a random variable. Hence the IV estimator is inconsistent when relevance fails. Moreover, the ratio of two independent normal random variables is Cauchy distributed, which has extremely heavy tails and no finite mean. This is why conventional normal approximations can fail badly.

5.2 Weak IV

Previously, we considered complete identification failure. Now suppose instead that identification is weak, in the sense that $\gamma = n^{-1/2}C$, where C is a constant. This local-to-zero formulation should not be taken literally; it is a useful asymptotic approximation. The parameter C indexes the strength of identification. As the sample size grows, the first-stage relationship becomes weaker at exactly the rate that keeps weak-instrument problems alive asymptotically.

Let us go back to

$$\hat{\beta}_{iv} - \beta = \frac{\sum_{i=1}^n Z_i e_i}{\sum_{i=1}^n Z_i X_i} = \frac{\widehat{Cov}(Z, e)}{\widehat{Cov}(Z, X)}$$

Asymptotically, the numerator tends to zero if the instrument is valid. But in finite samples it is never exactly zero, so if the denominator is also small, the ratio can be unstable. To see the problem more clearly, decompose the reduced-form error u into a part correlated with e and a part orthogonal to e:

$$u = \delta' e + \eta,$$

where $cov(\eta, e) = 0$ and $cov(z, \eta) = 0$. The parameter δ' controls the severity of the endogeneity problem. Now,

$$\hat{\beta}_{iv} - \beta = \frac{\widehat{Cov}(Z, e)}{\gamma\widehat{Var(Z)} + \widehat{Cov(Z, u)}}$$

Note $\widehat{Cov(Z, u)} = \delta' \widehat{Cov(Z, e)} + \widehat{Cov(Z, \eta)}$.

An instrument is weak if $\gamma Var(Z)$ is small enough that $\widehat{Cov}(Z,u)$ remains non-negligible relative to $\gamma\widehat{Var}(Z)$ even in large samples. Conversely, an instrument is strong if $\gamma Var(Z)$ is large enough that the sample covariance between Z and X is dominated by the true first-stage relationship rather than by spurious finite-sample covariance.

This is closely related to requiring the first-stage F-statistic to be large. The population analogue of the first-stage F-statistic is

$$F = n\frac{var(z\gamma)}{\sigma_u^2} = n\frac{\gamma^2\sigma_z^2}{\sigma_u^2}.$$

Note that $|\gamma\widehat{Var}(Z)| >> |\widehat{Cov}(Z,u)|$ can be written as

$$|\gamma|\widehat{Var}(Z) >> \hat{\sigma}_z\hat{\sigma}_u\widehat{corr}(z,u) \rightarrow \frac{|\gamma|\hat{\sigma}_z}{\hat{\sigma}_u} >> \widehat{corr}(z,u)$$

If Z is valid, $corr(z,u) = 0$, so $\widehat{corr}(z,u)$ converges to zero at $\sqrt{n}$ rate, and $\widehat{corr}(z,u)$ is bounded in probability by $\frac{k}{\sqrt{n}}$ for a positive constant $k > 0$. Thus we have

$$\frac{|\gamma|\hat{\sigma}_z}{\hat{\sigma}_u} >> \frac{k}{\sqrt{n}}$$

Finally, taking square, we get

$$\frac{|\gamma|\hat{\sigma}_z}{\hat{\sigma}_u} >> \frac{k}{\sqrt{n}} \rightarrow F = n\frac{\gamma^2\sigma_z^2}{\sigma_u^2} >> k^2$$

Thus, the intuitive requirement that the true first-stage signal dominate finite-sample noise translates into the requirement that the first-stage F be large.

We can show that the bias of $\hat{\beta}_{2sls}$ is approximately

$$\mathbb{E}\left[\hat{\beta}_{2sls} - \beta\right] \approx \frac{\sigma_{eu}}{\sigma_u^2}\frac{1}{F+1} = \delta\frac{1}{F+1}$$

So if F is small, the finite-sample bias can be substantial.

Moreover, under weak instruments the asymptotic distribution of the 2SLS estimator is non-normal, which makes the usual t-test unreliable. One important consequence is size distortion: if the instrument is weak and endogeneity is severe, a nominal 5% two-sided t-test of the true null H_0 : $\beta = 0$ may reject much more than 5% of the time.

How large does F need to be? The old rule of thumb was "first-stage $F > 10$." More recent work emphasizes that much larger values may be required for conventional inference to be accurate; For example "F-stat > 104.7" (Lee et al. 2022).

5.3 Anderson-Rubin test

There is a simple response to the poor behavior of the conventional 2SLS t-test: use the Anderson–Rubin (AR) test. To test $H_0 : \beta = \beta_0$, the AR test considers the regression

$$Y - \beta_0 X = \kappa Z + \xi$$

and tests $H_0 : \kappa = 0$. The intuition is simple: if H_0 is true, then the instrument Z should have no additional explanatory power for $Y - \beta_0 X$. In large samples, the AR statistic has a chi-squared approximation. When additional exogenous controls are present and $H_0 : \beta_0 = 0$, the AR test is equivalent to the test of the instrument in the reduced-form regression of Y on Z and the exogenous controls.

Because the regression and corresponding test statistic do not depend directly on the first-stage coefficient γ, the AR test is robust to weak instruments.

We can also invert the AR test to construct confidence intervals. Even when instruments are strong, it is often useful as a robustness check.

6 Local Average Treatment Effects

In this section, we introduce the design-based perspective on IV. In addition to drawing on the logic of experiments, the design-based framework also relaxes the constant-treatment-effect assumption commonly imposed in econometric models.

Suppose we are interested in the treatment effect of a binary variable D_i on Y_i. A canonical example is the causal effect of military service (D_i) on some later health outcome Y_i. The problem is that D_i is generally not independent of the potential outcomes $Y_i(0)$ and $Y_i(1)$. Fortunately, natural experiments sometimes generate exogenous variation in treatment take-up. During the Vietnam draft lottery, for example, lottery numbers created quasi-random variation in draft eligibility. Let Z_i denote draft status. If compliance with the draft had been perfect, then all those with a low lottery number ($Z = 1$) would have served in the military ($D = 1$), and all those with a high lottery number ($Z = 0$) would not have served ($D = 0$).

We can think of Z_i as random assignment and of $D_i(z)$ as the treatment actually taken when assignment is z. Thus treatment status itself has potential outcomes. There are four possible compliance types:

- A: Always Taker, if $D_i(1) = 1$ and $D_i(0) = 1$
- C: Complier, if $D_i(1) = 1$ and $D_i(0) = 0$
- D: Defier, if $D_i(1) = 0$ and $D_i(0) = 1$
- N: Never Taker, if $D_i(1) = 0$ and $D_i(0) = 0$

The observed $D_i = Z_i D_i(1) + (1 - Z_i)D_i(0)$. The potential outcome is a function of both Z_i and D_i: $Y_i(d, z)$. The observed outcome is

$$Y_i = (D_i Z_i)Y_i(1,1) + (D_i(1 - Z_i))Y_i(1,0) + ((1 - D_i)Z_i)Y_i(0,1) + ((1 - D_i)(1 - Z_i))Y_i(0,0)$$

6.1 Intention to treat effect

Because Z_i is random, we have

Assumption 4. *$Z_i \perp \{D_i(1), D_i(0), Y_i(d,z)\}$ for all d,z*

Therefore, we can identify the average causal effect of Z_i, which we call ITT (Intention to treat effect).

$$\begin{aligned}\mathbb{E}[Y_i|Z_i=1] - \mathbb{E}[Y_i|Z_i=0] &= \mathbb{E}[Y_i(D_i(1),1)|Z_i=1] - \mathbb{E}[Y_i(D_i(0),0)|Z_i=0] \\ &= \mathbb{E}[Y_i(D_i(1),1) - Y_i(D_i(0),0)]\end{aligned}$$

Although ITT is identified, it may not be the causal quantity we really care. We hope to know the causal effect of D_i.

6.2 Proportion of Compliers

We can also identify the average effect of Z_i on D_i.

$$\begin{aligned}\mathbb{E}[D_i|Z_i=1] - \mathbb{E}[D_i|Z_i=0] &= \mathbb{E}[D_i(1)|Z_i=1] - \mathbb{E}[D_i(0)|Z_i=0] \\ &= \mathbb{E}[D_i(1) - D_i(0)]\end{aligned}$$

If this quantity equals 1, then compliance is perfect. If it is less than 1, there is some noncompliance. In fact, we can learn more by decomposing $\tau_D = \mathbb{E}[D_i(1) - D_i(0)]$ by compliance type:

$$\begin{aligned}\tau_D =& \mathbb{E}[D_i(1) - D_i(0)|A]P(A) \quad \text{is 0 for A} \\ &+ \mathbb{E}[D_i(1) - D_i(0)|C]P(C) \quad \text{is 1 for C} \\ &+ \mathbb{E}[D_i(1) - D_i(0)|D]P(D) \quad \text{is -1 for D} \\ &+ \mathbb{E}[D_i(1) - D_i(0)|N]P(N) \quad \text{is 0 for N} \\ &= \mathbb{E}[D_i(1) - D_i(0)|C]P(C) + \mathbb{E}[D_i(1) - D_i(0)|D]P(D)\end{aligned}$$

Assumption 5 (Monotonicity)**.** $D_i(1) \geq D_i(0)$

This assumption says that no one defies their assignment; in other words, there are no defiers. It holds automatically under one-sided noncompliance when units assigned to control cannot access the treatment, so $D_i(0) = 0$ for all units. Therefore,

$$\begin{aligned}\tau_D &= \mathbb{E}[D_i(1) - D_i(0)|C]P(C) \\ \mathbb{E}[D_i|Z_i=1] - \mathbb{E}[D_i|Z_i=0] &= P(C) \quad \text{under assumption 4}\end{aligned}$$

In other words, we can identify the proportion of compliers.

6.3 LATE

Assumption 6 (Exclusion Restriction)**.** *$Y_i(d,0) = Y_i(d,1) = Y_i(d)$ for d=0,1*

Under this assumption, the potential outcome depends only on treatment status D, not directly on assignment Z.

We want to show that

$$\frac{\mathbb{E}\left[Y_i|Z_i=1\right]-\mathbb{E}\left[Y_i|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]}=\mathbb{E}\left[Y_i(1)-Y_i(0)|D_i(1)-D_i(0)=1\right].$$

The right-hand side is the causal effect of D_i on Y_i for units whose treatment status changes when assignment changes. Those are precisely the compliers. Thus the IV estimand identifies a *local average treatment effect* (LATE).

The denominator is now clear. Let us turn to the numerator. Under Assumption 4, the intention-to-treat effect is identified as

$$\tau_{ITT}=\mathbb{E}\left[Y_i|Z_i=1\right]-\mathbb{E}\left[Y_i|Z_i=0\right]=\mathbb{E}\left[Y_i(D_i(1),1)-Y_i(D_i(0),0)\right].$$

Under Assumption 6, this becomes

$$\tau_{ITT}=\mathbb{E}\left[Y_i|Z_i=1\right]-\mathbb{E}\left[Y_i|Z_i=0\right]=\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))\right].$$

As before, we can decompose it by compliance type:

$$\begin{aligned}\tau_{ITT}=&\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|A\right]P(A)\\&+\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|C\right]P(C)\\&+\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|D\right]P(D)\\&+\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|N\right]P(N)\end{aligned}$$

Based on the definition of four types, we know

$$\begin{aligned}\tau_{ITT}=&\mathbb{E}\left[Y_i(1)-Y_i(1)|A\right]P(A)\quad\text{is 0}\\&+\mathbb{E}\left[Y_i(1)-Y_i(0)|C\right]P(C)\\&+\mathbb{E}\left[Y_i(0)-Y_i(1)|D\right]P(D)\\&+\mathbb{E}\left[Y_i(0)-Y_i(0)|N\right]P(N)\quad\text{is 0}\\&=\mathbb{E}\left[Y_i(1)-Y_i(0)|C\right]P(C)+\mathbb{E}\left[Y_i(0)-Y_i(1)|D\right]P(D)\end{aligned}$$

Applying the no-defiers assumption 5, we obtain

$$\tau_{ITT}=E[Y_i|Z_i=1]-\mathbb{E}\left[Y_i|Z_i=0\right]=\mathbb{E}\left[Y_i(1)-Y_i(0)|C\right]P(C)$$

Recall that we already showed $P(C)=\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]$. Therefore,

$$E[Y_i|Z_i=1]-\mathbb{E}\left[Y_i|Z_i=0\right]=\mathbb{E}\left[Y_i(1)-Y_i(0)|C\right](\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right])$$

Assumption 7 (Relevance)**.** $\mathbb{E}\left[D_i(1)-D_i(0)\right]\neq 0$.

With this relevance assumption, we can divide by the first-stage effect and obtain

$$\begin{aligned}\frac{\mathbb{E}\left[Y_i|Z_i=1\right]-\mathbb{E}\left[Y_i|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]}&=\mathbb{E}\left[Y_i(1)-Y_i(0)|C\right]\\&=\mathbb{E}\left[Y_i(1)-Y_i(0)|D_i(1)-D_i(0)=1\right]\ \textbf{by assumption 5}\end{aligned}$$

Our proof and illustration is slightly different from Angrist et al. (1996).

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444–455.

Keane, M. P. & Neal, T. (2024). A practical guide to weak instruments. *Annual Review of Economics*, 16(1), 185–212.