Chapter 4. Instrumental Variables

Joan Llull

Quantitative Statistical Methods II
Barcelona School of Economics

Identification of Causal Effects in IV Settings

IV Settings

Suppose:

$$(Y_{1i}, Y_{0i}) \not\perp\!\!\!\perp D_i | X_i,$$

but there is some variable Z_i that satisfies **independence condition**:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i | X_i,$$

and the **relevance condition**:

$$Z_i \not\perp\!\!\!\perp D_i | X_i.$$

Matching can be regarded as a special case in which $Z_i = D_i$, i.e. all the variation in D_i is exogenous given X_i.

For simplicity, we do most of the analysis below considering a **single binary instrument** Z_i, and we abstract from including **covariates**.

Two cases: homogeneous and heterogeneous treatment effects.

Homogeneous Treatment Effects

In this case, the causal effect is the **same for every individual**:

$$Y_{1i} - Y_{0i} = \alpha \quad \forall i.$$

Availability of an instrumental variable allows us to **identify** α (IV regression):

$$\alpha = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, D_i)}.$$

Alternatively:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = Y_{0i} + \alpha D_i.$$

Taking into account that $Y_{0i} \perp\!\!\!\perp Z_i$ (conditional independence and subtract):

$$\alpha = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]},$$

Identification **requires** $\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] \neq 0$ (relevance).

We get the effect of D_i on Y_i through the effect of Z_i because Z_i only affects Y_i through D_i (**exclusion restriction**).

Heterogeneous Treatment Effects

In the heterogeneous case, the availability of instrumental variables is **not sufficient** to identify a causal effect (e.g. α_{ATE}).

Monotonicity condition: any person that was willing to treat if assigned to the control group would also be prepared to treat if assigned to the treatment group.

The **plausibility** of this assumption depends on the context of the application.

Under monotonicity, the IV coefficient coincides with the average treatment effect for those whose value of D_i would change when changing the value of Z_i, which is known as the **local average treatment effect** (LATE).

Potential Treatment Representation:

Let:

- D_{0i}: D_i when $Z_i = 0$
- D_{1i}: D_i when $Z_i = 1$

Only observe $D_{\ell i}$, for ℓ either equal to one or to zero $\Rightarrow$ **four observable** groups, eight **potential groups**:

Obs. type	Z_i	D_i	D_{0i}	D_{1i}	**Latent type**
Type 1	0	0	0	0	Never-taker
				1	Complier
Type 2	0	1	1	0	Defier
				1	Always-taker
Type 3	1	0	0	0	Never-taker
			1		Defier
Type 4	1	1	0	1	Complier
			1		Always-taker

Role of monotonicity:

Now we have:

$$\mathbb{E}[Y_i|Z_i = 1] = \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_{1i}]$$
$$\mathbb{E}[Y_i|Z_i = 0] = \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_{0i}],$$

which implies:

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] \\ &= \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = 1]P(D_{1i} - D_{0i} = 1) \\ &\quad - \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = -1]P(D_{1i} - D_{0i} = -1).\end{aligned}$$

Thus, $\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$ **could be negative** and yet the causal effect be **positive for everyone**, as long as the probability of *defiers* is sufficiently large.

Imperfect Compliance and IV

Imperfect Compliance and IV

Stronger assumption than monotonicity:

$$P(D_i = 1|Z_i = 0) = 0,$$

(no treatment for individuals with $Z_i = 0$) $\Rightarrow$ **imperfect compliance**.

In this case:

$$\mathbb{E}[Y_i|Z_i = 1] = \mathbb{E}[Y_{0i}] + \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1, Z_i = 1]P(D_i = 1|Z_i = 1),$$

and, since $P(D_i = 1|Z_i = 0) = 0$:

$$\mathbb{E}[Y_i|Z_i = 0] = \mathbb{E}[Y_{0i}].$$

Therefore:

$$\alpha_{TT} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{P(D_i = 1|Z_i = 1)}$$

(where we use $P(Z_i = 1|D_i = 1) = 1$.

Eligibility condition $\Rightarrow$ **IV coefficient** coincides with the **treatment effect on the treated**.

Local Average Treatment Effects (LATE)

Local Average Treatment Effects (LATE)

Ruling out defiers (which implies monotonicity):

$$\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]=\mathbb{E}[Y_{1i}-Y_{0i}|D_{1i}-D_{0i}=1]P(D_{1i}-D_{0i}=1),$$
$$\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]=\mathbb{E}[D_{1i}-D_{0i}]=P(D_{1i}-D_{0i}=1).$$

Local average treatment effect (LATE):

$$\alpha_{LATE}\equiv\mathbb{E}[Y_{1i}-Y_{0i}|D_{1i}-D_{0i}=1]=\frac{\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]}.$$

Imbens and Angrist (1994) called like this because it is the average treatment effects **on the subsample of compliers**.

$\Rightarrow$ different instrumental variables lead to **different parameters**, even under instrument validity, which is counter to standard GMM thinking.

$\Rightarrow$ importance of **group of compliers selected** by the instrument (policy relevance).

This concept changed radically the **way we think** of and understand IV.

Relevance requires presence of compliers.

Conditional Estimation with IV

Conditional Estimation with IV

Assume independence and relevance only cold **conditionally**:

$$\begin{array}{cc} (Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i | X_i & \text{(conditional independence)} \\ Z_i \not\perp\!\!\!\perp D_i | X_i & \text{(conditional relevance)} \ . \end{array}$$

Example: distance to college, Z_i, is not randomly assigned but chosen by parents, and this choice may depend on family background, X_i.

In general, we now have a **conditional LATE** given X_i:

$$\gamma(X_i) \equiv \mathbb{E}[Y_{1i} - Y_{0i} | D_{1i} - D_{0i} = 1, X_i],$$

and a conditional IV estimator:

$$\beta(X_i) \equiv \frac{\mathbb{E}[Y_i | Z_i = 1, X_i] - \mathbb{E}[Y_i | Z_i = 0, X_i]}{\mathbb{E}[D_i | Z_i = 1, X_i] - \mathbb{E}[D_i | Z_i = 0, X_i]}.$$

Aggregate effect: we proceed differently depending on whether the effects are homogeneous or heterogeneous.

In the **homogeneous** case: $Y_{1i} - Y_{0i} = \beta(X_i) \quad \forall i$.

In the **heterogeneous** case, it makes sense to consider an average treatment effect for the **overall subpopulation of compliers**:

$$\begin{aligned}\beta_C &\equiv \int \beta(X_i)\frac{P(compliers|X_i)}{P(compliers)}dF(X_i) \\ &= \int \{\mathbb{E}[Y_i|Z_i=1,X_i]-\mathbb{E}[Y_i|Z_i=0,X_i]\}\,\frac{1}{P(compliers)}dF(X_i),\end{aligned}$$

where:

$$P(compliers) = \int \{\mathbb{E}[D_i|Z_i=1,X_i]-\mathbb{E}[D_i|Z_i=0,X_i]\}\,dF(X_i).$$

Therefore:

$$\beta_C = \frac{\int \{\mathbb{E}[Y_i|Z_i=1,X_i]-\mathbb{E}[Y_i|Z_i=0,X_i]\}\,dF(X_i)}{\int \{\mathbb{E}[D_i|Z_i=1,X_i]-\mathbb{E}[D_i|Z_i=0,X_i]\}\,dF(X_i)},$$

which can be estimated as a **ratio of matching estimators** (Frölich, 2003).

Continuous Instruments: Marginal Treatment Effects (MTE)

Marginal Treatment Effects

Support of Z_i not binary: **multiplicity** of causal effects.

Which of these causal effects are **relevant** for evaluating a given policy?

We can define a different LATE parameter for **every pair** (z, z'):

$$\alpha_{LATE}(z, z') \equiv \frac{\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z']}{\mathbb{E}[D_i|Z_i = z] - \mathbb{E}[D_i|Z_i = z']}.$$

Multiplicity is even higher when there is **more than one** instrument.

Potential treatment status: as many potential treatment status indicators D_{zi} as possible values z of the instrument.

IV assumptions:

- Relevance: $P(D_i = 1|Z_i = z) \equiv P(z)$ is a nontrivial function of z.
- Orthogonality: $(Y_{0i}, Y_{1i}, D_{zi}) \perp\!\!\!\perp Z_i$.
- Monotonicity: $D_{zi} \geq D_{z'i}$ or $D_{zi} \leq D_{z'i}$ for any (z, z') and $\forall$ units in the population.

Using the **propensity score** $P(Z_i) \equiv P(D_i = 1|Z_i)$ as an instrument:

$$\alpha_{LATE}(P(z), P(z')) = \frac{\mathbb{E}[Y_i|P(Z_i) = P(z)] - \mathbb{E}[Y_i|P(Z_i) = P(z')]}{P(z) - P(z')}.$$

- **Binary** $Z_i \Rightarrow$ what we had in the first place
- **Continuous** $Z_i \Rightarrow$ marginal treatment effect (taking limits as $z \to z'$):

$$\alpha_{MTE}(P(z)) = \frac{\partial \mathbb{E}[Y_i|P(Z_i) = P(z)]}{\partial P(z)}.$$

$\alpha_{LATE}(P(z), P(z'))$ gives the ATE for individuals who would change schooling status from changing $P(Z_i)$ from $P(z)$ to $P(z')$.

Similarly, $\alpha_{MTE}(P(z))$ gives the ATE for individuals who would change treatment following a marginal change in $P(z)$, i.e. **indifferent between treatment choices** at $P(Z_i) = P(z)$.

Integrating $\alpha_{MTE}(U)$ over different ranges of U we can get other ATE measures. For example:

$$\alpha_{LATE}(P(z), P(z')) = \frac{\int_{P(z')}^{P(z)} \alpha_{MTE}(u)du}{P(z) - P(z')}, \text{ and } \alpha_{ATE} = \int_0^1 \alpha_{MTE}(u)du,$$

which makes it clear that to be able to identify α_{ATE} we need identification of $\alpha_{MTE}(u)$ over the **entire** $(0, 1)$ **range**.

Constructing suitably integrated marginal treatment effects, it may be possible to identify **policy relevant treatment effects.**

Remarks about Unobserved Heterogeneity in IV

Remarks about Unobserved Heterog. in IV

How important is it?

Balance between observed and unobserved heterogeneity depends on how detailed info on agents is available (empirical issue).

The worry is not heterogeneity *per se*, but the fact that heterogeneous gains may affect program participation.

Warnings:

In the absence of an economic model or a clear notional experiment, it is often difficult to interpret what IV estimates estimate.

Knowing that IV estimates can be interpreted as averages of heterogeneous effects is not very useful if understanding the heterogeneity itself is first order.

Heterogeneity of gains vs heterogeneity of treatments:

Heterogeneity of treatments may be more important. For example, the literature has found significant differences in returns to different college majors.

Problem of aggregating educational categ. is that returns are less meaningful.

Sometimes aggregated into just two categories, because some techniques are only well developed for binary explanatory variables.

Weak Instruments

Weak Instruments

If a relevant and orthogonal instrument Z_i is only **weakly correlated** with the treatment variable D_i we may have two complications:

- **Low precision**
- **Weak instruments bias**

The **low precision** problem is easy to see with one regressor, one instrument, and homoskedasticity. In that case:

$$\text{Var}(\hat{\beta}^{OLS}) = \frac{1}{N}\frac{\text{Var}(U_i)}{\text{Var}(D_i)} \leq \frac{1}{N}\frac{\text{Var}(U_i)}{\text{Var}(D_i)}\frac{1}{\rho^2_{D_i Z_i}} = \text{Var}(\hat{\beta}^{IV}),$$

where $\rho_{D_i Z_i} \equiv \text{Cov}(D_i, Z_i)/\sqrt{\text{Var}(D_i)\,\text{Var}(Z_i)}$ is the correlation coefficient between D_i and Z_i.

For the weak instruments bias, take the following **model**:

$$Y_i = \beta D_i + U_i$$
$$D_i = \pi Z_i + V_i.$$

The **IV coefficient** is:

$$\hat{\beta}^{IV} = \frac{\sum_{i=1}^N Z_i Y_i}{\sum_{i=1}^N Z_i D_i} = \beta + \frac{\sum_{i=1}^N Z_i U_i}{\sum_{i=1}^N \pi Z_i^2 + \sum_{i=1}^N Z_i V_i}.$$

The **case in which** $\pi \to 0$ is illustrative. In this case:

$$\hat{\beta}^{OLS} - \beta = \frac{\sum_{i=1}^N (\pi Z_i + V_i) U_i}{\sum_{i=1}^N (\pi Z_i + V_i)^2} \to \frac{\sum_{i=1}^N V_i U_i}{\sum_{i=1}^N V_i^2} \equiv \frac{\hat{\sigma}_{UV}}{\hat{\sigma}_V^2}.$$

and:

$$(\hat{\beta}^{IV} - \beta) \to \frac{\sum_{i=1}^N Z_i \left(\frac{\sigma_{UV}}{\sigma_V^2} V_i + \varepsilon_i \right)}{\sum_{i=1}^N Z_i V_i} = \frac{\sigma_{UV}}{\sigma_V^2} + \frac{\sum_{i=1}^N Z_i \varepsilon_i}{\sum_{i=1}^N Z_i V_i}.$$

Assuming the **expectation** exists:

$$\mathbb{E}[\hat{\beta}^{IV} - \beta] \to \frac{\sigma_{UV}}{\sigma_V^2} + \mathbb{E}\left[\frac{\sum_{i=1}^N Z_i \varepsilon_i}{\sum_{i=1}^N Z_i V_i}\right] = \frac{\sigma_{UV}}{\sigma_V^2}.$$

Thus, when the instrument is weak, there is an IV sample bias that **tends to the OLS bias** as the instrument tends to irrelevant.

With more tedious algebra (available in the book) we obtain, for $\pi \neq 0$, the following expression:

$$\mathbb{E}[\hat{\beta}^{IV} - \beta] \approx \frac{\sigma_{UV}}{\sigma_V} \frac{1}{\mathbb{E}[F]},$$

where $\mathbb{E}[F] \equiv \frac{\mathbb{E}[\pi' Z' Z \pi]/K}{\sigma_V^2} + 1$, the expectation of the F-**statistic** for the test of the coefficients on **excluded instruments** in the first-stage regression equal to zero.

$F > 0$ by construction $\Rightarrow$ **sign** weak instruments bias = **sign** OLS bias.

Rule of thumb: $F > 10 \Rightarrow$ strong instruments. But, **in reality**, the level of concern depends on numerous factors: size of the OLS bias, number of excluded instruments, number of observations,....