Lecture 10: Instrumental variables and linear GMM

Economics 527: Econometric Methods

Vadim Marmer, UBC

Endogeneity

Endogenous and exogenous regressors

- Regression, $k = k_1 + k_2$:

$$Y_i = \underset{1\times k_1}{X_{i,1}^\top}\ \underset{k_1\times 1}{\beta_1} + \underset{1\times k_2}{X_{i,2}^\top}\ \underset{k_2\times 1}{\beta_2} + U_i$$

- $X_{i,1}$ endogenous:

$$\mathrm{E}\left[X_{i,1}U_i\right] \neq 0$$

- $X_{i,2}$ exogenous:

$$\mathrm{E}\left[X_{i,2}U_i\right] = 0$$

Omitted variables

- Log wage on education; $X_{i,2}$: exogenous, for example gender:

$$\log \mathrm{Wage}_i = \beta_1 \,\mathrm{Education}_i + X_{i,2}^\top \beta_2 + U_i$$

- Ability: unobserved, in the error; ε_i: luck:

$$U_i = \mathrm{ability}_i + \varepsilon_i$$

- Education chosen optimally by i given ability:

$$\mathrm{E}\left[\mathrm{Education}_i \times \mathrm{ability}_i\right] \neq 0 \quad \Longrightarrow \quad \mathrm{Education}_i \ \text{ endogenous}$$

Simultaneity: demand and supply

- Market i, demand and supply:

$$\underset{1\times 1}{Y_i^d} = \beta_1 P_i + \beta_2 I_i + U_i^d, \qquad \beta_1 < 0$$
$$Y_i^s = \gamma_1 P_i + \gamma_2 G_i + U_i^s, \qquad \gamma_1 > 0$$

- I_i: observed demand shifter (e.g. income).
- G_i: observed supply shifter (e.g. input prices, weather).
- U_i^d, U_i^s: unobserved demand and supply shifters.
- Y_i^d, Y_i^s: not observed; observed in equilibrium:

$$Y_i = Y_i^d = Y_i^s, \qquad P_i: \text{ equilibrium price}$$

Equilibrium price

- Supply equals demand; solve for P_i:

$$\begin{aligned}\underbrace{\gamma_1 P_i + \gamma_2 G_i + U_i^s}_{Y_i^s} &= \underbrace{\beta_1 P_i + \beta_2 I_i + U_i^d}_{Y_i^d} \\ (\gamma_1 - \beta_1) P_i &= -\gamma_2 G_i + \beta_2 I_i + U_i^d - U_i^s \\ P_i &= \underbrace{\frac{-\gamma_2}{\gamma_1 - \beta_1}}_{\pi_1} G_i + \underbrace{\frac{\beta_2}{\gamma_1 - \beta_1}}_{\pi_2} I_i + \frac{U_i^d - U_i^s}{\gamma_1 - \beta_1}\end{aligned}$$

- Common denominator in P_i:

$$\gamma_1 - \beta_1 > 0$$

Price and the demand error

- Demand regression:

$$Y_i = \beta_1 P_i + \beta_2 I_i + U_i^d$$

- **Assume:** U_i^d, U_i^s uncorrelated with G_i, I_i and with each other:

$$\mathrm{E}\left[G_i U_i^d\right] = 0, \qquad \mathrm{E}\left[I_i U_i^d\right] = 0, \qquad \mathrm{E}\left[U_i^s U_i^d\right] = 0$$

- Substitute the equilibrium price:

$$\begin{aligned}\mathrm{E}\left[P_i U_i^d\right] &= \mathrm{E}\left[\left(\pi_1 G_i + \pi_2 I_i + \frac{U_i^d - U_i^s}{\gamma_1 - \beta_1}\right) U_i^d\right] \\ &= \mathrm{E}\left[\pi_1 G_i U_i^d + \pi_2 I_i U_i^d + \frac{(U_i^d)^2 - U_i^s U_i^d}{\gamma_1 - \beta_1}\right] \\ &= \pi_1 \underbrace{\mathrm{E}\left[G_i U_i^d\right]}_{=0} + \pi_2 \underbrace{\mathrm{E}\left[I_i U_i^d\right]}_{=0} + \frac{\mathrm{E}\left[(U_i^d)^2\right] - \overbrace{\mathrm{E}\left[U_i^s U_i^d\right]}^{=0}}{\gamma_1 - \beta_1} \\ &= \frac{\mathrm{E}\left[(U_i^d)^2\right]}{\gamma_1 - \beta_1} > 0\end{aligned}$$

- P_i: endogenous in the demand equation.

OLS under endogeneity

OLS estimator of β_1

- One endogenous regressor, $k_1 = 1$:

$$Y_i = \underset{1\times 1}{\beta_1}\ \underset{1\times 1}{X_{i,1}} + X_{i,2}^\top \beta_2 + U_i, \qquad \mathrm{E}\left[X_{i,1} U_i\right] \neq 0, \qquad \mathrm{E}\left[X_{i,2} U_i\right] = 0$$

- Recall, partitioned regression, $M_2 = I_n - X_2 (X_2^\top X_2)^{-1} X_2^\top$:

$$\begin{aligned}\hat{\beta}_1 &= \underset{n\times 1}{(X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y} \\ &= \frac{X_1^\top M_2 Y}{X_1^\top M_2 X_1} \\ &= \beta_1 + \frac{\frac{1}{n} X_1^\top M_2 U}{\frac{1}{n} X_1^\top M_2 X_1} \qquad (M_2 X_2 = 0)\end{aligned}$$

Denominator of $\hat{\beta}_1 - \beta_1$

- **Assume:** (Y_i, X_i) iid; $\mathrm{E}\left[\|X_i\|^2\right] < \infty$; $\mathrm{E}\left[U_i^2\right] < \infty$.
- Recall: weak LLN for sample averages, Slutsky for products and inverses.
- Expand M_2, take limits:

$$\begin{aligned}
\underbrace{\frac{1}{n} X_1^\top M_2 X_1}_{1\times 1} &= \frac{X_1^\top X_1}{n} - \frac{X_1^\top X_2}{n}\Big(\frac{X_2^\top X_2}{n}\Big)^{-1}\frac{X_2^\top X_1}{n} \\
&\to_p \underbrace{\mathrm{E}\left[X_{i,1}^2\right]}_{Q_{11}} - \underbrace{\mathrm{E}\left[X_{i,1}X_{i,2}^\top\right]}_{Q_{12}} \underbrace{\left(\mathrm{E}\left[X_{i,2}X_{i,2}^\top\right]\right)^{-1}}_{Q_{22}^{-1}} \underbrace{\mathrm{E}\left[X_{i,2}X_{i,1}\right]}_{Q_{21}} \\
&= Q_{1\cdot 2} > 0 \qquad \left(\mathrm{E}\left[X_i X_i^\top\right] \text{ p.d.}\right)
\end{aligned}$$

Numerator of $\hat{\beta}_1 - \beta_1$

- Weak LLN; expand M_2, take limits:

$$\begin{aligned}
\frac{X_1^\top U}{n} &= \frac{1}{n}\sum_{i=1}^n X_{i,1}U_i \\
&\to_p \mathrm{E}\left[X_{i,1}U_i\right] \\
\frac{1}{n}X_1^\top M_2 U &= \frac{X_1^\top U}{n} - \frac{X_1^\top X_2}{n}\Big(\frac{X_2^\top X_2}{n}\Big)^{-1}\frac{X_2^\top U}{n} \\
&\to_p \mathrm{E}\left[X_{i,1}U_i\right] - Q_{12}\,Q_{22}^{-1}\underbrace{\mathrm{E}\left[X_{i,2}U_i\right]}_{=0} \\
&= \mathrm{E}\left[X_{i,1}U_i\right] \neq 0
\end{aligned}$$

Asymptotic bias of $\hat{\beta}_1$

- Slutsky:

$$\begin{aligned}
\hat{\beta}_1 &= \beta_1 + \frac{\frac{1}{n}X_1^\top M_2 U}{\frac{1}{n}X_1^\top M_2 X_1} \\
&\to_p \beta_1 + \frac{\overbrace{\mathrm{E}\left[X_{i,1}U_i\right]}^{\neq 0}}{Q_{1\cdot 2}} \neq \beta_1
\end{aligned}$$

- Asymptotic bias of OLS:

$$\boxed{\mathrm{AsyBias}(\hat{\beta}_1) = \frac{\mathrm{E}\left[X_{i,1}U_i\right]}{Q_{1\cdot 2}}}$$

- Wage example, $X_{i,1}$: education, U_i: ability:

$$\mathrm{E}\left[X_{i,1}U_i\right] > 0 \quad \Longrightarrow \quad \mathrm{AsyBias}(\hat{\beta}_1) > 0$$

- $\hat{\beta}_1$ over-estimates β_1.

OLS estimator of β_2

- Coefficients on the exogenous regressors, $M_1 = I_n - X_1(X_1^\top X_1)^{-1}X_1^\top$:

$$\begin{aligned}
\hat{\beta}_2 &= (X_2^\top M_1 X_2)^{-1} X_2^\top M_1 Y \\
&= \beta_2 + \Big(\frac{X_2^\top M_1 X_2}{n}\Big)^{-1}\frac{X_2^\top M_1 U}{n}
\end{aligned}$$

- First factor, inside the inverse:

$$\frac{X_2^\top M_1 X_2}{n} \to_p Q_{2\cdot 1} = Q_{22} - Q_{21}\, Q_{11}^{-1} Q_{12} \qquad \text{p.d.}$$

- Second factor:

$$\begin{aligned}\frac{X_2^\top M_1 U}{n} = & \underbrace{\frac{X_2^\top U}{n}}_{\to_p \mathrm{E}[X_{i,2}U_i]=0} - \underbrace{\frac{X_2^\top X_1}{n}}_{\to_p Q_{21}} \underbrace{\left(\frac{X_1^\top X_1}{n}\right)^{-1}}_{\to_p Q_{11}^{-1}} \underbrace{\frac{X_1^\top U}{n}}_{\to_p \mathrm{E}[X_{i,1}U_i]\neq 0} \\ & \to_p -Q_{21}\, Q_{11}^{-1}\, \mathrm{E}\left[X_{i,1}U_i\right]\end{aligned}$$

Asymptotic bias of $\hat{\beta}_2$

- Slutsky; scalar $Q_{11} = \mathrm{E}\left[X_{i,1}^2\right]$:

$$\begin{aligned}\hat{\beta}_2 &\to_p \beta_2 - \underbrace{Q_{2\cdot 1}^{-1}}_{\text{p.d.}}\, Q_{21}\, Q_{11}^{-1}\, \mathrm{E}\left[X_{i,1}U_i\right] \\ &= \beta_2 - \frac{Q_{2\cdot 1}^{-1}}{Q_{11}}\, Q_{21}\, \mathrm{E}\left[X_{i,1}U_i\right]\end{aligned}$$

- $\hat{\beta}_2$: asymptotically biased unless

$$Q_{21} = \mathrm{E}\left[X_{i,2}X_{i,1}\right] = 0$$

Example: education and gender

- Wage equation, U_i: ability:

$$\log \mathrm{Wage}_i = \beta_1 \underbrace{\mathrm{Education}_i}_{X_{i,1}} + \beta_2 \underbrace{\mathrm{Female}_i}_{X_{i,2},\ \text{dummy}} + U_i$$

- Suppose:

$$\mathrm{E}\left[X_{i,1}U_i\right] > 0, \qquad Q_{21} = \mathrm{E}\left[\mathrm{Female}_i\, \mathrm{Education}_i\right] > 0$$

- Asymptotic bias of $\hat{\beta}_2$:

$$\hat{\beta}_2 - \beta_2 \to_p -\underbrace{Q_{2\cdot 1}^{-1}}_{+}\ \underbrace{Q_{21}}_{+}\ \underbrace{Q_{11}^{-1}}_{+}\ \underbrace{\mathrm{E}\left[X_{i,1}U_i\right]}_{+} < 0$$

Identification and instruments

Identification without endogeneity

- No endogeneity:

$$Y_i = X_i^\top \beta + U_i, \qquad \beta \in \mathbb{R}^k, \qquad \mathrm{E}\left[X_i U_i\right] = 0$$

- Moment condition; k equations in k unknowns; $\mathrm{E}\left[X_i X_i^\top\right]$ invertible:

$$\begin{aligned}& \mathrm{E}\left[X_i \overbrace{(Y_i - X_i^\top \beta)}^{U_i}\right] = 0 \\ \iff\quad & \underbrace{\mathrm{E}\left[X_i Y_i\right]}_{k\times 1} = \underbrace{\mathrm{E}\left[X_i X_i^\top\right]}_{k\times k}\ \underbrace{\beta}_{k\times 1} \\ \implies\quad & \beta = \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1} \mathrm{E}\left[X_i Y_i\right] \qquad \text{(identification)}\end{aligned}$$

- Both moments estimated by sample averages:

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top \to_p \mathrm{E}\left[X_i X_i^\top\right]$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \to_p \mathrm{E}\left[X_i Y_i\right]$$

Identification with endogeneity

- $X_{i,1}$ endogenous: k_1 equations lost:

$$\mathrm{E}\left[X_{i,1} U_i\right] \neq 0, \qquad \mathrm{E}\left[X_{i,2} U_i\right] = 0$$

- Remaining moment condition:

$$\mathrm{E}\left[\underset{k_2\times 1}{X_{i,2}}(Y_i - \underset{k_1\times 1}{X_{i,1}^\top \beta_1} - \underset{k_2\times 1}{X_{i,2}^\top \beta_2})\right] = 0$$

- k_2 equations, $k_1 + k_2$ unknowns $\Longrightarrow$ no unique solution for β_1, β_2: not identified.

Instrumental variables (IVs)

- Suppose we also observe $Z_{i,1}$, an $\ell_1 \times 1$ vector of variables such that:

$$\mathrm{E}\left[Z_{i,1} U_i\right] = 0 \quad \Longrightarrow \quad \ell_1 \text{ additional equations}$$

- Two sets of moment conditions:

$$\mathrm{E}\left[X_{i,2}(Y_i - X_{i,1}^\top \beta_1 - X_{i,2}^\top \beta_2)\right] = 0 \qquad (k_2 \text{ equations})$$
$$\mathrm{E}\left[Z_{i,1}(Y_i - X_{i,1}^\top \beta_1 - X_{i,2}^\top \beta_2)\right] = 0 \qquad (\ell_1 \text{ equations})$$

Stacked moment condition

- Regressors X_i and exogenous variables (IVs) Z_i:

$$X_i = \begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} : \ \underbrace{(k_1 + k_2)}_{=k} \times 1$$
$$Z_i = \begin{pmatrix} Z_{i,1} \\ X_{i,2} \end{pmatrix} : \ \underbrace{(\ell_1 + k_2)}_{=\ell} \times 1$$
$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

- Moment condition:

$$\mathrm{E}\left[Z_i(Y_i - X_i^\top \beta)\right] = 0$$
$$\iff \quad \underbrace{\mathrm{E}\left[Z_i X_i^\top\right]}_{\ell\times k} \underset{k\times 1}{\beta} = \underbrace{\mathrm{E}\left[Z_i Y_i\right]}_{\ell\times 1}$$

- $\ell = \ell_1 + k_2$ equations, $k = k_1 + k_2$ unknowns.

One regressor, one instrument

- Simplified version, Y_i, X_i, U_i, Z_i all 1×1:

$$Y_i = \beta X_i + U_i$$

- Covariance with the instrument:

$$\begin{aligned} \operatorname{Cov}(Z_i, Y_i) &= \beta \operatorname{Cov}(Z_i, X_i) + \overbrace{\operatorname{Cov}(Z_i, U_i)}^{=0} \\ \implies \quad \beta &= \frac{\operatorname{Cov}(Z_i, Y_i)}{\operatorname{Cov}(Z_i, X_i)} \qquad \text{if } \operatorname{Cov}(Z_i, X_i) \neq 0 \end{aligned}$$

Exactly identified case

- $\ell = k$ ($\ell_1 = k_1$): k equations in k unknowns; $\mathrm{E}[Z_i X_i^\top]$: $k \times k$, rank k:

$$\begin{aligned} 0 &= \mathrm{E}[Z_i(Y_i - X_i^\top \beta)] \\ &= \underbrace{\mathrm{E}[Z_i Y_i]}_{\text{in place of } \operatorname{Cov}(Z_i, Y_i)} - \underbrace{\mathrm{E}[Z_i X_i^\top]}_{\text{in place of } \operatorname{Cov}(Z_i, X_i)} \beta \\ \implies \quad \beta &= \boxed{(\mathrm{E}[Z_i X_i^\top])^{-1} \mathrm{E}[Z_i Y_i]} \end{aligned}$$

IV estimator

- Z: $n \times \ell$, with rows $Z_i^\top$, as X with rows $X_i^\top$:

$$Z = \begin{pmatrix} Z_1^\top \\ \vdots \\ Z_n^\top \end{pmatrix}$$

$$\sum_{i=1}^n Z_i X_i^\top = Z^\top X$$

$$\sum_{i=1}^n Z_i Y_i = Z^\top Y$$

- Sample averages in place of expectations:

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

Over-identified case

- $\ell > k$ ($\ell_1 > k_1$), more equations than unknowns; $\mathrm{E}[Z_i X_i^\top]$: $\ell \times k$.
- Premultiply by $\mathrm{E}[X_i Z_i^\top]$, which is $k \times \ell$:

$$\begin{aligned} \mathrm{E}[Z_i X_i^\top]\,\beta &= \mathrm{E}[Z_i Y_i] \\ \implies \quad \underbrace{(\mathrm{E}[X_i Z_i^\top])(\mathrm{E}[Z_i X_i^\top])}_{k \times k} \beta &= (\mathrm{E}[X_i Z_i^\top])\mathrm{E}[Z_i Y_i] \\ \implies \quad \beta &= \Big[(\mathrm{E}[X_i Z_i^\top])(\mathrm{E}[Z_i X_i^\top])\Big]^{-1} (\mathrm{E}[X_i Z_i^\top])\mathrm{E}[Z_i Y_i] \end{aligned}$$

- Condition for identification: $k \times k$ inverse exists.

Rank condition

- The $\ell \times k$ matrix Q:

$$Q = \mathrm{E}\left[Z_i X_i^\top\right], \qquad Q^\top = \mathrm{E}\left[X_i Z_i^\top\right]$$

- **Rank condition.**

$$\operatorname{rank}(Q) = \operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k$$

- $\operatorname{rank}(Q) = k \Longrightarrow Q^\top Q$ p.d.; for $c \in \mathbb{R}^k$ and $v = Qc$:

$$\begin{aligned} c^\top Q^\top Q c &= v^\top v \\ &= \|v\|^2 \\ &> 0 \quad \text{unless } c = 0 \qquad (v = Qc \neq 0 \ \text{ unless } c = 0) \end{aligned}$$

- Rank condition: necessary and sufficient for identification of β.

Order condition

- Necessary for the rank condition:

$$\begin{aligned} \operatorname{rank}\underset{\ell\times k}{\left(Q\right)} = k \quad &\Longrightarrow \quad \ell \geq k \\ &\Longleftrightarrow \quad \ell_1 + k_2 \geq k_1 + k_2 \\ &\Longleftrightarrow \quad \underbrace{\ell_1}_{\text{number of excluded IVs}} \geq \underbrace{k_1}_{\text{number of endogenous regressors}} \end{aligned}$$

- Three cases:

$$\begin{aligned} \operatorname{rank}(Q) = k,\ \ell = k &: \text{ exactly identified} \\ \operatorname{rank}(Q) = k,\ \ell > k &: \text{ over-identified} \\ \operatorname{rank}(Q) < k &: \text{ under-identified} \quad (\ell < k, \text{ or linear dependence}) \end{aligned}$$

Included regressor as instrument

- Two scalar regressors:

$$Y_i = \beta_1 X_{i,1} + \beta_2 X_{i,2} + U_i, \qquad \mathrm{E}\left[X_{i,2} U_i\right] = 0$$

- Try $Z_{i,1} = X_{i,2}$:

$$\begin{aligned} \underset{2\times 2}{\mathrm{E}\left[Z_i X_i^\top\right]} &= \mathrm{E}\left[\begin{pmatrix} Z_{i,1} \\ X_{i,2} \end{pmatrix} \begin{pmatrix} X_{i,1} & X_{i,2} \end{pmatrix}\right] \\ &= \begin{pmatrix} \mathrm{E}\left[Z_{i,1} X_{i,1}\right] & \mathrm{E}\left[Z_{i,1} X_{i,2}\right] \\ \mathrm{E}\left[X_{i,2} X_{i,1}\right] & \mathrm{E}\left[X_{i,2} X_{i,2}\right] \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{E}\left[X_{i,2} X_{i,1}\right] & \mathrm{E}\left[X_{i,2} X_{i,2}\right] \\ \mathrm{E}\left[X_{i,2} X_{i,1}\right] & \mathrm{E}\left[X_{i,2} X_{i,2}\right] \end{pmatrix} \qquad (Z_{i,1} = X_{i,2}) \end{aligned}$$

- Two equal rows $\Longrightarrow$ rank $1 < k = 2$.
- **Exclusion restriction:** $Z_{i,1}$ excluded from the equation for Y_i; no direct effect on Y_i.

Noise as instrument

- Suppose $Z_{i,1}$ is "noise", with $\mathrm{E}\left[Z_{i,1}U_i\right] = 0$:

$$\mathrm{E}\left[Z_i X_i^\top\right] = \begin{pmatrix} \overbrace{\mathrm{E}\left[Z_{i,1}X_{i,1}\right]}^{=0} & \overbrace{\mathrm{E}\left[Z_{i,1}X_{i,2}\right]}^{=0} \\ \mathrm{E}\left[X_{i,2}X_{i,1}\right] & \mathrm{E}\left[X_{i,2}X_{i,2}\right] \end{pmatrix}$$

- Row of zeros $\implies$ rank $1 < k = 2$.

First stage and the rank condition

First stage

- First stage, for the endogenous regressors:

$$\underset{k_1\times 1}{X_{i,1}} = \underset{k_1\times \ell_1}{\Pi_1^\top}\ \underset{\ell_1\times 1}{Z_{i,1}} + \underset{k_1\times k_2}{\Pi_2^\top}\ \underset{k_2\times 1}{X_{i,2}} + \underset{k_1\times 1}{V_i}$$

- V_i uncorrelated with $Z_{i,1}$ and $X_{i,2}$:

$$\mathrm{E}\left[Z_{i,1}V_i^\top\right] = 0, \qquad \mathrm{E}\left[X_{i,2}V_i^\top\right] = 0$$

- First stage in terms of Z_i, with Π: $\ell \times k_1$:

$$X_{i,1} = \Pi^\top Z_i + V_i, \qquad \Pi = \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix}, \qquad \mathrm{E}\left[Z_i V_i^\top\right] = 0$$

Definition of Π

- **Claim.** With $\mathrm{E}\left[Z_i Z_i^\top\right]$ p.d., Π can always be defined so that $\mathrm{E}\left[Z_i V_i^\top\right] = 0$.
- Transpose the first stage, premultiply by Z_i, take expectations:

$$\begin{aligned} X_{i,1}^\top &= Z_i^\top \Pi + V_i^\top \\ Z_i X_{i,1}^\top &= Z_i Z_i^\top \Pi + Z_i V_i^\top \\ \mathrm{E}\left[Z_i X_{i,1}^\top\right] &= \mathrm{E}\left[Z_i Z_i^\top\right]\Pi + \underbrace{\mathrm{E}\left[Z_i V_i^\top\right]}_{=0} \\ \implies \quad \Pi &= \left(\mathrm{E}\left[Z_i Z_i^\top\right]\right)^{-1}\mathrm{E}\left[Z_i X_{i,1}^\top\right] \end{aligned}$$

- Given $X_{i,1}$ and Z_i, define:

$$\begin{aligned} \Pi &\equiv \left(\mathrm{E}\left[Z_i Z_i^\top\right]\right)^{-1}\mathrm{E}\left[Z_i X_{i,1}^\top\right] \\ V_i &\equiv X_{i,1} - \Pi^\top Z_i \end{aligned}$$

Proof that Z_i and V_i are uncorrelated

- With Π and V_i as defined:

$$\begin{aligned} \mathrm{E}\left[Z_i V_i^\top\right] &= \mathrm{E}\left[Z_i(X_{i,1}^\top - Z_i^\top \Pi)\right] \\ &= \mathrm{E}\left[Z_i X_{i,1}^\top\right] - \mathrm{E}\left[Z_i Z_i^\top\right]\Pi \\ &= \mathrm{E}\left[Z_i X_{i,1}^\top\right] - \mathrm{E}\left[Z_i Z_i^\top\right]\left(\mathrm{E}\left[Z_i Z_i^\top\right]\right)^{-1}\mathrm{E}\left[Z_i X_{i,1}^\top\right] \\ &= 0 \end{aligned}$$

- Hence the first stage can always be defined:

$$X_{i,1} = \Pi^\top Z_i + V_i, \qquad \mathrm{E}\left[Z_i V_i^\top\right] = 0$$

Rank condition in terms of Π_1

- **Claim.** $\mathrm{E}\left[Z_i Z_i^\top\right]$ p.d.; Π_1: $\ell_1 \times k_1$ first-stage coefficients on $Z_{i,1}$:

$$\operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k \quad \iff \quad \operatorname{rank}\left(\Pi_1\right) = k_1$$

Proof, step 1: $\mathrm{E}\left[Z_i X_i^\top\right]$ in blocks

- We will use the transposed first stage, with $\mathrm{E}\left[Z_{i,1} V_i^\top\right] = 0$, $\mathrm{E}\left[X_{i,2} V_i^\top\right] = 0$:

$$X_{i,1}^\top = Z_{i,1}^\top \Pi_1 + X_{i,2}^\top \Pi_2 + V_i^\top$$

- Write in blocks; substitute the first stage:

$$\begin{aligned}
\mathrm{E}\left[Z_i X_i^\top\right] &= \mathrm{E}\left[\begin{pmatrix} Z_{i,1} \\ X_{i,2} \end{pmatrix} \begin{pmatrix} X_{i,1}^\top & X_{i,2}^\top \end{pmatrix}\right] \\
&= \begin{pmatrix} \mathrm{E}\left[Z_{i,1} X_{i,1}^\top\right] & \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \\ \mathrm{E}\left[X_{i,2} X_{i,1}^\top\right] & \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{E}\left[Z_{i,1} Z_{i,1}^\top\right] \Pi_1 + \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \Pi_2 & \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \\ \mathrm{E}\left[X_{i,2} Z_{i,1}^\top\right] \Pi_1 + \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \Pi_2 & \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \end{pmatrix}
\end{aligned}$$

Proof, step 2: $\mathrm{E}\left[Z_i X_i^\top\right] \theta$

- Multiply by $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$, with θ_1: $k_1 \times 1$ and θ_2: $k_2 \times 1$:

$$\begin{aligned}
\mathrm{E}\left[Z_i X_i^\top\right] \theta &= \begin{pmatrix} (\mathrm{E}\left[Z_{i,1} Z_{i,1}^\top\right] \Pi_1 + \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \Pi_2)\theta_1 + \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \theta_2 \\ (\mathrm{E}\left[X_{i,2} Z_{i,1}^\top\right] \Pi_1 + \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \Pi_2)\theta_1 + \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \theta_2 \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{E}\left[Z_{i,1} Z_{i,1}^\top\right] \Pi_1 \theta_1 + \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] (\Pi_2 \theta_1 + \theta_2) \\ \mathrm{E}\left[X_{i,2} Z_{i,1}^\top\right] \Pi_1 \theta_1 + \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] (\Pi_2 \theta_1 + \theta_2) \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} \mathrm{E}\left[Z_{i,1} Z_{i,1}^\top\right] & \mathrm{E}\left[Z_{i,1} X_{i,2}^\top\right] \\ \mathrm{E}\left[X_{i,2} Z_{i,1}^\top\right] & \mathrm{E}\left[X_{i,2} X_{i,2}^\top\right] \end{pmatrix}}_{= \mathrm{E}[Z_i Z_i^\top],\ \ell \times \ell,\ \text{p.d.}} \begin{pmatrix} \Pi_1 \theta_1 \\ \Pi_2 \theta_1 + \theta_2 \end{pmatrix}
\end{aligned}$$

Proof, step 3: sufficiency of $\operatorname{rank}\left(\Pi_1\right) = k_1$

- $\mathrm{E}\left[Z_i Z_i^\top\right]$ p.d.; suppose $\operatorname{rank}\left(\Pi_1\right) = k_1$:

$$\begin{aligned}
\mathrm{E}\left[Z_i X_i^\top\right] \theta = 0 \quad &\iff \quad \Pi_1 \theta_1 = 0 \ \text{ and } \ \Pi_2 \theta_1 + \theta_2 = 0 \\
&\iff \quad \theta_1 = 0 \ \text{ and } \ \Pi_2 \theta_1 + \theta_2 = 0 \qquad (\operatorname{rank}\left(\Pi_1\right) = k_1) \\
&\iff \quad \theta_1 = 0 \ \text{ and } \ \theta_2 = 0 \\
&\iff \quad \theta = 0
\end{aligned}$$

- k columns of $\mathrm{E}\left[Z_i X_i^\top\right]$ linearly independent:

$$\operatorname{rank}\left(\Pi_1\right) = k_1 \quad \implies \quad \operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k$$

Proof, step 4: necessity of $\operatorname{rank}\left(\Pi_1\right) = k_1$

- Suppose $\operatorname{rank}\left(\Pi_1\right) < k_1$:

$$\begin{aligned}
&\text{there is } \ \theta_1 \neq 0 \ \text{ such that } \ \Pi_1 \theta_1 = 0 \\
&\text{define } \ \theta_2 = -\Pi_2 \theta_1 \\
\implies \ & \Pi_2 \theta_1 + \theta_2 = 0 \\
\implies \ & \mathrm{E}\left[Z_i Z_i^\top\right] \begin{pmatrix} \Pi_1 \theta_1 \\ \Pi_2 \theta_1 + \theta_2 \end{pmatrix} = 0 \\
\implies \ & \mathrm{E}\left[Z_i X_i^\top\right] \theta = 0 \quad \text{with } \ \theta \neq 0 \\
\implies \ & \operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) < k
\end{aligned}$$

- Hence:

$$\operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k \quad \Longrightarrow \quad \operatorname{rank}\left(\Pi_1\right) = k_1$$

One endogenous regressor

- Suppose $k_1 = 1$, for example $Y_i = \log \text{Wage}_i$ and $X_{i,1} = \text{Education}_i$:

$$\underset{1\times 1}{\text{Education}_i} = \underset{1\times \ell_1}{\Pi_1^\top}\ \underset{\ell_1 \times 1}{Z_{i,1}} + \Pi_2^\top X_{i,2} + V_i, \qquad \Pi_1 = \begin{pmatrix} \Pi_{1,1} \\ \vdots \\ \Pi_{1,\ell_1} \end{pmatrix}$$

- First stage written instrument by instrument:

$$X_{i,1} = \sum_{j=1}^{\ell_1} Z_{i,1,j}\,\Pi_{1,j} + \Pi_2^\top X_{i,2} + V_i$$

- Rank condition:

$$\begin{aligned} \operatorname{rank}\left(\Pi_1\right) = 1 \quad &\Longleftrightarrow \quad \Pi_1 \neq 0 \\ &\Longleftrightarrow \quad \Pi_{1,j} \neq 0 \ \text{ for at least one } j = 1, \ldots, \ell_1 \end{aligned}$$

Requirements for instruments

- We need $Z_{i,1}$: $\ell_1 \times 1$, with $\ell_1 \geq k_1$, such that:
  1. $Z_{i,1}$ excluded from the equation for Y_i.
  2. $Z_{i,1}$ exogenous: $\mathrm{E}\left[Z_{i,1} U_i\right] = 0$.
  3. $Z_{i,1}$ informative about $X_{i,1}$: $\operatorname{rank}\left(\Pi_1\right) = k_1$ in the first stage.
- Wage example, U_i: ability; IV must be:
  1. Independent of ability.
  2. Not affecting the wage directly.
  3. Shifting education.

Examples of instruments

- Angrist and Krueger (1991), IVs: quarter-of-birth dummies:

$$\log \text{Wage}_i = \beta_1 \,\text{Education}_i + X_{i,2}^\top \beta_2 + U_i$$

- Angrist and Evans (1998), married women with at least 2 children:

$$\begin{aligned} \text{Hours worked}_i &= \beta_1 \,(\text{number of children})_i + X_{i,2}^\top \beta_2 + U_i \\ (\text{number of children})_i &= \gamma_1 \,\text{Hours worked}_i + \ldots \end{aligned}$$

- IV: same-sex dummy.

Generalized method of moments (GMM)

Weight matrix W

- **Definition.** Weight matrix W: $\ell \times \ell$, symmetric, p.d.
- Premultiply by $Q^\top W$:

$$\begin{aligned} & Q\,\beta = \mathrm{E}\left[Z_i Y_i\right] \\ \Longrightarrow \quad & \underbrace{\underset{k\times \ell}{Q^\top}\ \underset{\ell \times \ell}{W}\ \underset{\ell \times k}{Q}}_{k \times k}\ \beta = Q^\top W\,\mathrm{E}\left[Z_i Y_i\right] \\ & \Longrightarrow \quad \beta = \left(Q^\top W Q\right)^{-1} Q^\top W\,\mathrm{E}\left[Z_i Y_i\right] \end{aligned}$$

- $W = I_\ell \Longrightarrow \beta = \left(Q^\top Q\right)^{-1} Q^\top \mathrm{E}\left[Z_i Y_i\right]$.

Exactly identified case: W cancels

- $\ell = k$; Q, W: $k \times k$, invertible:

$$\begin{aligned}\beta &= (\underset{k\times k}{Q^\top}\ \underset{k\times k}{W}\ \underset{k\times k}{Q}\ )^{-1}\ \underset{k\times k}{Q^\top}\ \underset{k\times k}{W}\ \mathrm{E}\left[Z_i Y_i\right] \\ &= Q^{-1}\, W^{-1}(Q^\top)^{-1} Q^\top W\, \mathrm{E}\left[Z_i Y_i\right] \\ &= Q^{-1}\, \mathrm{E}\left[Z_i Y_i\right]\end{aligned}$$

- Sample averages in place of expectations:

$$\begin{aligned}(X^\top Z\, W\, Z^\top X)^{-1} X^\top Z\, W\, Z^\top Y &= (Z^\top X)^{-1} Z^\top Y \\ &= \hat{\beta}^{IV}\end{aligned}$$

GMM estimator $\hat{\beta}(W_n)$

- **Definition.** GMM estimator with weight matrix W_n:

$$\hat{\beta}(W_n) = (X^\top Z\, W_n\, Z^\top X)^{-1} X^\top Z\, W_n\, Z^\top Y$$

- **Assumption I.** $\{(Y_i, X_i, Z_i) : i = 1, \ldots, n\}$ iid; $Y_i = X_i^\top \beta + U_i$, $\beta \in \mathbb{R}^k$.
- **Assumption II.** $\mathrm{E}\left[Z_i U_i\right] = 0$.
- **Assumption III.** $\mathrm{rank}\,(Q) = k$, $Q = \mathrm{E}\left[Z_i X_i^\top\right]$.
- **Assumption IV.** $W_n \to_p W$; W: $\ell \times \ell$, symmetric, p.d.

Consistency of $\hat{\beta}(W_n)$

- Weak LLN, Assumptions I and II:

$$\begin{aligned}\frac{Z^\top X}{n} &= \frac{1}{n}\sum_{i=1}^n Z_i X_i^\top \to_p \mathrm{E}\left[Z_i X_i^\top\right] = Q \\ \frac{Z^\top U}{n} &= \frac{1}{n}\sum_{i=1}^n Z_i U_i \to_p \mathrm{E}\left[Z_i U_i\right] = 0\end{aligned}$$

- Slutsky, Assumptions III and IV:

$$\left(\frac{X^\top Z}{n}\, W_n\, \frac{Z^\top X}{n}\right)^{-1} \to_p (Q^\top W Q)^{-1}$$

- **Theorem.** Under Assumptions I to IV:

$$\begin{aligned}\hat{\beta}(W_n) &= \beta + (X^\top Z\, W_n\, Z^\top X)^{-1} X^\top Z\, W_n\, Z^\top U \qquad (Y = X\beta + U) \\ &= \beta + \underbrace{\left(\frac{X^\top Z}{n}\, W_n\, \frac{Z^\top X}{n}\right)^{-1}}_{\to_p (Q^\top W Q)^{-1}}\ \underbrace{\frac{X^\top Z}{n}\, W_n}_{\to_p Q^\top W}\ \underbrace{\frac{Z^\top U}{n}}_{\to_p 0} \\ &\to_p \beta\end{aligned}$$

CLT for $Z^\top U/\sqrt{n}$

- **Assumption V.** $\Omega = \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right]$ finite and p.d.
- Variance of $Z_i U_i$, with $\mathrm{E}\left[Z_i U_i\right] = 0$:

$$\begin{aligned}\Omega &= \mathrm{Var}\,(Z_i U_i) \\ &= \mathrm{E}\left[Z_i U_i (Z_i U_i)^\top\right] \\ &= \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right]\end{aligned}$$

- $Z_1U_1, \ldots, Z_nU_n$: iid; multivariate CLT:

$$\frac{Z^\top U}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i U_i \to_d N(0, \Omega)$$

Asymptotic variance of $\hat{\beta}(W_n)$

- Recall, Cramér convergence theorem: $A_n \to_p A$, A non-random, $B_n \to_d B \implies A_n B_n \to_d AB$.
- Multiply the representation of $\hat{\beta}(W_n) - \beta$ by $\sqrt{n}$:

$$\begin{aligned}
\sqrt{n}(\hat{\beta}(W_n) - \beta) &= \underbrace{\Big(\frac{X^\top Z}{n} W_n \frac{Z^\top X}{n}\Big)^{-1}}_{\to_p (Q^\top W Q)^{-1}} \underbrace{\frac{X^\top Z}{n} W_n}_{\to_p Q^\top W} \underbrace{\frac{Z^\top U}{\sqrt{n}}}_{\to_d N(0,\Omega)} \\
&\to_d (Q^\top W Q)^{-1} Q^\top W \, N(0, \Omega) \\
&\stackrel{d}{=} N(0, \underbrace{(Q^\top W Q)^{-1} Q^\top W \Omega W Q (Q^\top W Q)^{-1}}_{V(W)}) \qquad (W, \ Q^\top W Q \text{ symmetric})
\end{aligned}$$

- **Theorem.** Under Assumptions I to V:

$$\sqrt{n}(\hat{\beta}(W_n) - \beta) \to_d N(0, V(W)), \qquad V(W) = \text{AsyVar}(\hat{\beta}(W_n))$$

Efficient weight matrix $W^* = \Omega^{-1}$

- **Definition.** Efficient weight matrix W^*: minimizes $V(W)$ in the p.s.d. sense:

$$V(W) - V(W^*) \geq 0 \quad \text{(p.s.d.) for every } W$$

- **Claim.** $W^* = \Omega^{-1}$.
- $V(W)$ at $W = \Omega^{-1}$:

$$\begin{aligned}
V(\Omega^{-1}) &= (Q^\top \Omega^{-1} Q)^{-1} \underbrace{Q^\top \Omega^{-1} \Omega \, \Omega^{-1} Q}_{= Q^\top \Omega^{-1} Q} (Q^\top \Omega^{-1} Q)^{-1} \\
&= (Q^\top \Omega^{-1} Q)^{-1}
\end{aligned}$$

- $W^* = \Omega^{-1}$ matters only if $\ell > k$; $\ell = k \implies W$ cancels.

Comparing asymptotic variances

- Scalars, $a, b > 0$:

$$a - b \geq 0 \quad \iff \quad \frac{1}{b} - \frac{1}{a} \geq 0$$

- Matrices, p.s.d. differences:

$$\begin{aligned}
& V(W) - V(\Omega^{-1}) \geq 0 \\
\iff \quad & \big(V(\Omega^{-1})\big)^{-1} - \big(V(W)\big)^{-1} \geq 0
\end{aligned}$$

Proof that $W^* = \Omega^{-1}$ is efficient

- Ω p.d. $\implies \Omega^{1/2}$: symmetric, $\Omega^{1/2}\,\Omega^{1/2} = \Omega$; $\Omega^{-1/2} = (\Omega^{1/2})^{-1}$.
- Recall: $M_H = I - H(H^\top H)^{-1} H^\top$: symmetric, idempotent, p.s.d.

- Difference of the inverses:

$$\begin{aligned}(V(\Omega^{-1}))^{-1} - (V(W))^{-1} &= Q^\top \Omega^{-1} Q - (Q^\top W Q)(Q^\top W \Omega W Q)^{-1}(Q^\top W Q) \\ &= Q^\top \Omega^{-1/2}\Big(I - \underbrace{\Omega^{1/2} W Q}_{H}\,(\underbrace{Q^\top W \Omega W Q}_{H^\top H})^{-1}\,\underbrace{Q^\top W \Omega^{1/2}}_{H^\top}\Big)\Omega^{-1/2} Q \\ &= Q^\top \Omega^{-1/2}\ \underbrace{M_H}_{\text{p.s.d.}}\ \Omega^{-1/2} Q \\ &\geq 0\end{aligned}$$

Two-step efficient GMM

- Efficient weight matrix, unknown:

$$W^* = \Omega^{-1} = \left(\mathrm{E}\left[U_i^2 Z_i Z_i^\top\right]\right)^{-1}$$

- Step 1: $W_n = I_\ell$; $\hat{\beta}(I_\ell)$ consistent; estimator of Ω from its residuals:

$$\hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i^\top \hat{\beta}(I_\ell)\right)^2 Z_i Z_i^\top$$

- Step 2: $W_n = \hat{\Omega}_n^{-1}$, efficient GMM:

$$\hat{\beta}(\hat{\Omega}_n^{-1}) = (X^\top Z\, \hat{\Omega}_n^{-1} Z^\top X)^{-1} X^\top Z\, \hat{\Omega}_n^{-1} Z^\top Y$$

- One can show: $\hat{\Omega}_n \to_p \Omega$.

Asymptotic variance of $\hat{\beta}^{IV}$

- $\ell = k$: Q: $k \times k$, invertible; IV estimator:

$$\begin{aligned}\hat{\beta}^{IV} &= (\underbrace{Z^\top X}_{k\times k})^{-1} Z^\top Y \\ &= \beta + (Z^\top X)^{-1} Z^\top U \qquad (Y = X\beta + U)\end{aligned}$$

- Cramér convergence theorem:

$$\begin{aligned}\sqrt{n}(\hat{\beta}^{IV} - \beta) &= \underbrace{\left(\frac{Z^\top X}{n}\right)^{-1}}_{\to_p Q^{-1}}\ \underbrace{\frac{Z^\top U}{\sqrt{n}}}_{\to_d N(0,\Omega)} \\ &\to_d Q^{-1}\, N(0, \Omega)\end{aligned}$$

- Variance of $Q^{-1}N(0,\Omega)$; $(Q^{-1})^\top = (Q^\top)^{-1}$:

$$\begin{aligned}\mathrm{AsyVar}(\hat{\beta}^{IV}) &= Q^{-1}\,\Omega\,(Q^{-1})^\top \\ &= \boxed{\left(\mathrm{E}\left[Z_i X_i^\top\right]\right)^{-1} \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right] \left(\mathrm{E}\left[X_i Z_i^\top\right]\right)^{-1}}\end{aligned}$$

Two-stage least squares (2SLS)

Homoskedasticity: efficient GMM is 2SLS

- Suppose:

$$\mathrm{E}\left[U_i^2 \mid Z_i\right] = \sigma^2 \quad \text{(homoskedasticity)}$$

- Law of iterated expectations (LIE):

$$\begin{aligned}\Omega &= \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right] \\ &\overset{\text{LIE}}{=} \sigma^2 \,\mathrm{E}\left[Z_i Z_i^\top\right]\end{aligned}$$

- Estimator of Ω:

$$\hat{\Omega}_n = \hat{\sigma}_n^2 \frac{Z^\top Z}{n}$$

- Efficient GMM; the scalar $n/\hat{\sigma}_n^2$ cancels:

$$\begin{aligned}\hat{\beta}(\hat{\Omega}_n^{-1}) &= \left(X^\top Z\left(\hat{\sigma}_n^2 \frac{Z^\top Z}{n}\right)^{-1} Z^\top X\right)^{-1} X^\top Z\left(\hat{\sigma}_n^2 \frac{Z^\top Z}{n}\right)^{-1} Z^\top Y \\ &= \left(X^\top \underbrace{Z(Z^\top Z)^{-1} Z^\top}_{P_Z} X\right)^{-1} X^\top \underbrace{Z(Z^\top Z)^{-1} Z^\top}_{P_Z} Y \\ &= \boxed{(X^\top P_Z X)^{-1} X^\top P_Z Y \equiv \hat{\beta}^{2SLS}}\end{aligned}$$

2SLS: first and second stage

- First stage, OLS of $X_{i,1}$ on Z_i:

$$\begin{aligned}\hat{\Pi} &= (Z^\top Z)^{-1} Z^\top X_1 \\ \hat{X}_{i,1} &= \hat{\Pi}^\top Z_i \\ &= \hat{\Pi}_1^\top Z_{i,1} + \hat{\Pi}_2^\top X_{i,2}\end{aligned}$$

- Fitted values; X_2: columns of $Z \implies P_Z X_2 = X_2$:

$$\begin{aligned}\hat{X} &= P_Z X \\ \hat{X}_i &= \begin{pmatrix}\hat{X}_{i,1} \\ X_{i,2}\end{pmatrix}\end{aligned}$$

- Second stage, OLS of Y on $\hat{X}$; $P_Z = P_Z^\top P_Z$:

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

Control function and the Hausman test

Every regressor endogenous

- $k_2 = 0$: $X_i = X_{i,1}$, $Z_i = Z_{i,1}$.
- Model and first stage, Π: $\ell \times k$:

$$\begin{aligned}Y_i &= X_i^\top \beta + U_i \\ \underset{k\times 1}{X_i} &= \underset{k\times \ell}{\Pi^\top}\, \underset{\ell \times 1}{Z_i} + \underset{k\times 1}{V_i}\end{aligned}$$

- Exogenous instruments; Π defined so that Z_i, V_i uncorrelated:

$$\mathrm{E}\left[Z_i U_i\right] = 0, \qquad \mathrm{E}\left[Z_i V_i^\top\right] = 0$$

- V_i: the source of endogeneity:

$$\mathrm{E}\left[X_i U_i\right] \neq 0, \quad \mathrm{E}\left[Z_i U_i\right] = 0 \quad \implies \quad \mathrm{E}\left[V_i U_i\right] \neq 0$$

Decomposition of U_i

- **Define:**

$$\begin{aligned} \lambda &= \left(\mathrm{E}\left[V_i V_i^\top\right]\right)^{-1} \mathrm{E}\left[V_i U_i\right] \\ \varepsilon_i &= U_i - V_i^\top \lambda \\ \implies \quad U_i &= V_i^\top \lambda + \varepsilon_i \end{aligned}$$

- V_i and ε_i uncorrelated:

$$\begin{aligned} \mathrm{E}\left[V_i \varepsilon_i\right] &= \mathrm{E}\left[V_i U_i\right] - \mathrm{E}\left[V_i V_i^\top\right] \lambda \\ &= \mathrm{E}\left[V_i U_i\right] - \mathrm{E}\left[V_i V_i^\top\right] \left(\mathrm{E}\left[V_i V_i^\top\right]\right)^{-1} \mathrm{E}\left[V_i U_i\right] \\ &= 0 \end{aligned}$$

Augmented regression and control function

- Z_i and ε_i uncorrelated:

$$\begin{aligned} \mathrm{E}\left[Z_i \varepsilon_i\right] &= \mathrm{E}\left[Z_i (U_i - V_i^\top \lambda)\right] \\ &= \underbrace{\mathrm{E}\left[Z_i U_i\right]}_{=0} - \underbrace{\mathrm{E}\left[Z_i V_i^\top\right]}_{=0} \lambda \\ &= 0 \end{aligned}$$

- X_i and ε_i uncorrelated:

$$\begin{aligned} \mathrm{E}\left[X_i \varepsilon_i\right] &= \mathrm{E}\left[\left(\Pi^\top Z_i + V_i\right) \varepsilon_i\right] \\ &= \Pi^\top \underbrace{\mathrm{E}\left[Z_i \varepsilon_i\right]}_{=0} + \underbrace{\mathrm{E}\left[V_i \varepsilon_i\right]}_{=0} \\ &= 0 \end{aligned}$$

- Substitute $U_i = V_i^\top \lambda + \varepsilon_i$:

$$\begin{aligned} Y_i &= X_i^\top \beta + V_i^\top \lambda + \varepsilon_i \\ \mathrm{E}\left[X_i \varepsilon_i\right] &= 0 \\ \mathrm{E}\left[V_i \varepsilon_i\right] &= 0 \end{aligned}$$

- No endogeneity; V_i: **control function**.

Control function (CF) estimator

- First-stage residuals:

$$\hat{V}_i = X_i - \hat{\Pi}^\top Z_i, \qquad \hat{\Pi} = (Z^\top Z)^{-1} Z^\top X$$

- CF estimator, OLS of Y_i on X_i and $\hat{V}_i$:

$$Y_i = X_i^\top \hat{\beta}^{CF} + \hat{V}_i^\top \hat{\lambda} + \hat{\varepsilon}_i$$

- **Claim.** $\hat{\beta}^{CF} = \hat{\beta}^{2SLS}$.

Proof that $\hat{\beta}^{CF} = \hat{\beta}^{2SLS}$

- Matrix form; $Z\hat{\Pi} = P_Z X$:

$$\underset{n\times 1}{Y} = \underset{n\times k}{X}\hat{\beta}^{CF} + \hat{V}\hat{\lambda} + \hat{\varepsilon}, \qquad \hat{V} = X - Z\hat{\Pi} = M_Z X$$

- Recall, partitioned regression, $M_{\hat{V}} = I_n - \hat{V}(\hat{V}^\top \hat{V})^{-1}\hat{V}^\top$:

$$\hat{\beta}^{CF} = (X^\top M_{\hat{V}} X)^{-1} X^\top M_{\hat{V}} Y$$

- M_Z symmetric, idempotent $\Longrightarrow \hat{V}^\top \hat{V} = X^\top M_Z X$:

$$\begin{aligned} X^\top M_{\hat{V}} X &= X^\top X - X^\top \hat{V}(\hat{V}^\top \hat{V})^{-1}\hat{V}^\top X \\ &= X^\top X - X^\top M_Z X\,(X^\top M_Z X)^{-1} X^\top M_Z X \\ &= X^\top (I_n - M_Z) X \\ &= X^\top P_Z X \end{aligned}$$

- Same steps, with Y in place of the last X:

$$X^\top M_{\hat{V}} Y = X^\top P_Z Y$$

- Hence:

$$\boxed{\hat{\beta}^{CF} = (X^\top P_Z X)^{-1} X^\top P_Z Y = \hat{\beta}^{2SLS}}$$

Hausman test of exogeneity

- $\hat{\lambda}$: estimator of $\lambda = \left(\mathrm{E}\left[V_i V_i^\top\right]\right)^{-1}\mathrm{E}\left[V_i U_i\right]$.
- Exogeneity of the regressors in terms of λ:

$$\begin{aligned} \lambda = 0 \quad &\Longleftrightarrow \quad \mathrm{E}\left[V_i U_i\right] = 0 \\ &\Longleftrightarrow \quad \mathrm{E}\left[X_i U_i\right] = 0 \qquad \left(\mathrm{E}\left[X_i U_i\right] = \Pi^\top \mathrm{E}\left[Z_i U_i\right] + \mathrm{E}\left[V_i U_i\right]\right) \end{aligned}$$

- Hypotheses:

$$H_0 : \lambda = 0 \;\; (X_i \text{ exogenous}) \qquad \text{vs} \qquad H_1 : \lambda \neq 0 \;\; (X_i \text{ endogenous})$$

- Under H_0, one can show:

$$H_n = n\,\hat{\lambda}^\top \left[\widehat{\mathrm{Asy\,Var}}(\hat{\lambda})\right]^{-1}\hat{\lambda} \to_d \chi^2_{k_1}, \qquad k_1 : \text{ number of endogenous regressors}$$

- Reject H_0: $H_n > \chi^2_{k_1,\,1-\alpha}$.

Over-identifying restrictions test

Exactly identified case: $Z^\top \hat{U} = 0$

- Linear IV model, $\beta \in \mathbb{R}^k$, $Z_i \in \mathbb{R}^\ell$:

$$Y_i = X_i^\top \beta + U_i, \qquad \mathrm{E}\left[Z_i U_i\right] = 0, \qquad \operatorname{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k$$

- $\ell = k$, $\hat{U} = Y - X\hat{\beta}^{IV}$:

$$\begin{aligned} Z^\top \hat{U} &= Z^\top Y - Z^\top X \hat{\beta}^{IV} \\ &= Z^\top Y - Z^\top X (Z^\top X)^{-1} Z^\top Y \\ &= 0 \end{aligned}$$

Hypotheses and efficient GMM

- $\ell > k$:

$$H_0: \ \mathrm{E}\left[Z_i(Y_i - X_i^\top \beta)\right] = 0 \quad \text{for some } \beta \in \mathbb{R}^k$$
$$H_1: \ \mathrm{E}\left[Z_i(Y_i - X_i^\top b)\right] \neq 0 \quad \text{for all } b \in \mathbb{R}^k$$

- $\hat{\beta} = \hat{\beta}(\hat{\Omega}_n^{-1})$: efficient GMM, with $\hat{\Omega}_n \to_p \Omega$:

$$\Omega = \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right] = \mathrm{E}\left[(Y_i - X_i^\top \beta)^2 Z_i Z_i^\top\right]$$

Statistic J

- $\hat{\Omega}_n^{-1}$-weighted distance of $\frac{1}{n}\sum_{i=1}^n Z_i \hat{U}_i$ from zero, $\hat{U}_i = Y_i - X_i^\top \hat{\beta}$:

$$\begin{aligned} J &= \frac{1}{n}(Y - X\hat{\beta})^\top Z \ \hat{\Omega}_n^{-1} \ \frac{1}{n} Z^\top (Y - X\hat{\beta}) \\ &= \Big(\frac{1}{n}\sum_{i=1}^n Z_i \hat{U}_i\Big)^\top \hat{\Omega}_n^{-1} \Big(\frac{1}{n}\sum_{i=1}^n Z_i \hat{U}_i\Big) \end{aligned}$$

- $\ell = k$: $\hat{\beta} = \hat{\beta}^{IV}$, $Z^\top \hat{U} = 0 \implies J = 0$, whether H_0 is true or false.
- $\ell > k$, H_0 true:

$$\frac{1}{n}\sum_{i=1}^n Z_i \hat{U}_i \to_p 0 \quad \implies \quad J > 0, \text{ but } J \to_p 0$$

Quadratic form at the true β

- Under H_0, CLT:

$$\frac{1}{\sqrt{n}} Z^\top (Y - X\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i U_i \to_d N(0, \Omega)$$

- We will use, for $x \in \mathbb{R}^\ell$:

$$\begin{aligned} x^\top \Omega^{-1} x &= x^\top \Omega^{-1/2} \, \Omega^{-1/2} x \\ &= (\Omega^{-1/2} x)^\top \, \Omega^{-1/2} x \\ &= \|\Omega^{-1/2} x\|^2 \end{aligned}$$

- $\mathcal{Z} = \Omega^{-1/2} N(0, \Omega) \sim N(0, I_\ell)$:

$$\begin{aligned} \|\mathcal{Z}\|^2 &= \mathcal{Z}^\top \mathcal{Z} \\ &= \sum_{j=1}^\ell \mathcal{Z}_j^2 \sim \chi^2_\ell \end{aligned}$$

- $\hat{\Omega}_n^{-1} = \Omega^{-1} + o_p(1)$:

$$\begin{aligned} \frac{(Y - X\beta)^\top Z}{\sqrt{n}} \, \hat{\Omega}_n^{-1} \, \frac{Z^\top (Y - X\beta)}{\sqrt{n}} &= \frac{U^\top Z}{\sqrt{n}} \, \Omega^{-1} \underbrace{\frac{Z^\top U}{\sqrt{n}}}_{\to_d N(0,\Omega)} + o_p(1) \\ &= \Big\| \underbrace{\Omega^{-1/2} \frac{Z^\top U}{\sqrt{n}}}_{\to_d \mathcal{Z}} \Big\|^2 + o_p(1) \\ &\to_d \|\mathcal{Z}\|^2 \sim \chi^2_\ell \end{aligned}$$

J test and Sargan's test

- Under H_0, one can show, with $\widehat{\beta}$ efficient GMM:

$$nJ = \frac{(Y - X\widehat{\beta})^\top Z}{\sqrt{n}} \, \widehat{\Omega}_n^{-1} \, \frac{Z^\top (Y - X\widehat{\beta})}{\sqrt{n}} \to_d \chi^2_{\ell - k}$$

- k IVs used to estimate β, $\ell - k$ IVs used to test.
- Test: reject H_0 when $nJ > \chi^2_{\ell-k,\,1-\alpha}$.
- $\mathrm{E}\left[U_i^2 \mid Z_i\right] = \sigma^2 \implies \widehat{\Omega}_n = \widehat{\sigma}_n^2 \, \frac{Z^\top Z}{n}$, $\widehat{\beta} = \widehat{\beta}^{2SLS}$; J test: Sargan's test.

Constructing more instruments

Functions of exogenous variables as instruments

- Suppose:

$$\mathrm{E}\left[U_i \mid X_{i,2}, Z_{i,1}\right] = 0$$

- Law of iterated expectations, any function h:

$$\begin{aligned}
\mathrm{E}\left[h(X_{i,2}, Z_{i,1})\, U_i\right] &= \mathrm{E}\left[\mathrm{E}\left[h(X_{i,2}, Z_{i,1})\, U_i \mid X_{i,2}, Z_{i,1}\right]\right] \\
&= \mathrm{E}\left[h(X_{i,2}, Z_{i,1}) \underbrace{\mathrm{E}\left[U_i \mid X_{i,2}, Z_{i,1}\right]}_{=0}\right] \\
&= 0
\end{aligned}$$

- $Z_{i,1}$: IVs $\implies$ products of entries of $Z_{i,1}$ and $X_{i,2}$: IVs too.

Angrist and Krueger instruments

- $X_{i,2}$: gender, race, year-of-birth and geographic dummies.
- Quarter-of-birth dummies:

$$Z_{i,1} = \begin{pmatrix} \mathbf{1}(\text{quarter of birth}_i = 1) \\ \mathbf{1}(\text{quarter of birth}_i = 2) \\ \mathbf{1}(\text{quarter of birth}_i = 3) \end{pmatrix}$$

- More IVs: each quarter-of-birth dummy times each year-of-birth dummy.

Summary

Endogeneity and identification

- $\mathrm{E}\left[X_{i,1} U_i\right] \neq 0 \implies$ OLS inconsistent; $k_1 = 1$: $\mathrm{AsyBias}(\widehat{\beta}_1) = \mathrm{E}\left[X_{i,1} U_i\right] / Q_{1\cdot 2}$.
- $k_1 = 1$: $\widehat{\beta}_2$ also asymptotically biased unless $\mathrm{E}\left[X_{i,2} X_{i,1}\right] = 0$.
- $Z_i = (Z_{i,1}^\top, X_{i,2}^\top)^\top$, $\mathrm{E}\left[Z_i (Y_i - X_i^\top \beta)\right] = 0$: ℓ equations, k unknowns.
- Rank condition $\mathrm{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k \iff \mathrm{rank}\left(\Pi_1\right) = k_1$; order condition $\ell_1 \geq k_1$.
- $Z_{i,1}$: excluded, exogenous, informative; $\ell = k$: $\widehat{\beta}^{IV} = (Z^\top X)^{-1} Z^\top Y$.

GMM and 2SLS

- $\hat{\beta}(W_n) = (X^\top Z W_n Z^\top X)^{-1} X^\top Z W_n Z^\top Y$; Assumptions I to IV $\implies \hat{\beta}(W_n) \to_p \beta$.
- Assumptions I to V $\implies \sqrt{n}(\hat{\beta}(W_n) - \beta) \to_d N(0, V(W))$, $V(W) = (Q^\top W Q)^{-1} Q^\top W \Omega W Q (Q^\top W Q)^{-1}$.
- Efficient weight matrix: $W^* = \Omega^{-1}$, $V(\Omega^{-1}) = (Q^\top \Omega^{-1} Q)^{-1}$; two steps, $\hat{\Omega}_n$ from $\hat{\beta}(I_\ell)$.
- $\ell = k \implies \hat{\beta}(W_n) = \hat{\beta}^{IV}$; $\text{AsyVar}(\hat{\beta}^{IV}) = Q^{-1}\Omega(Q^{-1})^\top$.
- $\mathrm{E}\left[U_i^2 \mid Z_i\right] = \sigma^2 \implies$ efficient GMM $= \hat{\beta}^{2SLS} = (X^\top P_Z X)^{-1} X^\top P_Z Y = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top Y$, $\hat{X} = P_Z X$.

Control function and specification tests

- $U_i = V_i^\top \lambda + \varepsilon_i$: $Y_i = X_i^\top \beta + V_i^\top \lambda + \varepsilon_i$, $\mathrm{E}\left[X_i \varepsilon_i\right] = 0$, $\mathrm{E}\left[V_i \varepsilon_i\right] = 0$; $\hat{\beta}^{CF} = \hat{\beta}^{2SLS}$.
- Hausman: $H_0 : \lambda = 0$; $H_n \to_d \chi^2_{k_1}$ under H_0; reject when $H_n > \chi^2_{k_1,\,1-\alpha}$.
- Over-identifying restrictions, $\ell > k$: $nJ \to_d \chi^2_{\ell-k}$ under H_0; reject when $nJ > \chi^2_{\ell-k,\,1-\alpha}$; homoskedasticity $\implies$ Sargan's test.
- $\mathrm{E}\left[U_i \mid X_{i,2}, Z_{i,1}\right] = 0 \implies \mathrm{E}\left[h(X_{i,2}, Z_{i,1})\, U_i\right] = 0$: functions of $X_{i,2}$, $Z_{i,1}$ as IVs.