Core Econometrics - Endogeneity and IV

Table of Contents

MT Part III: Endogeneity and IV

3 1 Endogeneity

Measurement Errors

Measurement Error in Y $\rightsquigarrow$ No Problem #flashcard

- Consider an additive, zero-mean, uncorrelated with x_i measurement error in the dependent variable only:

$$y_i = y_i^\star + v_i \iff y_i^\star = y_i - v_i$$

  - y_i is the observed value
  - $y_i^\star$ is the true value
  - $\mathbb{E}[v_i] = 0$ (zero mean)
  - $\mathbb{E}[x_i u_i] = 0$ (important assumption)
- We estimate the model:

$$\begin{aligned} y_i &= y_i^\star + u_i \\ &= x_i\beta + v_i + u_i \end{aligned}$$

- We still have $\mathbb{E}[(v_i + u_i)x_i] = 0 \implies$ OLS is consistent

Measurement Error in X $\rightsquigarrow$ Attenuation Bias (Case of Classical Errors-in-variable Assumptions) #flashcard

- General setup:

$$x_i = x_i^\star + e_i \iff x_i^\star = x_i - e_i$$

- Classical Error-in-variable assumptions:
  - $\mathbb{E}[x_i^\star e_i] = 0$ measurement error is uncorrelated with the true value of $x_i^\star$
  - $\mathbb{E}[u_i e_i] = 0$ measurement error is uncorrelated with the true model error u_i
  - $Var[e_i] = \sigma_e^2$ measurement error is homoskedastic
  - $Var[x_i^\star] = \sigma_{x^\star}^2$ population variance of the true $x_i^\star$ exists and is finite
- SLR

Now $\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y^* = \dfrac{\sum_{i=1}^{n} x_i y_i^*}{\sum_{i=1}^{n} x_i^2} = \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i^*}{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$

Use scalars here

Using $x_i = x_i^* + e_i$ and $y_i^* = x_i^*\beta + u_i$ together with the above assumptions, we obtain

10

$$\underset{n\to\infty}{\text{p lim}}\ \widehat{\beta}_{\text{OLS}} = \frac{\text{p lim}\frac{1}{n}\sum_{i=1}^{n}(x_i^*+e_i)(x_i^*\beta+u_i)}{\text{p lim}\frac{1}{n}\sum_{i=1}^{n}(x_i^*+e_i)^2}$$

Slutsky Theorem

$$= \frac{\left(\text{p lim}\frac{1}{n}\sum_{i=1}^{n} x_i^{*2}\right)\beta + \text{p lim}\frac{1}{n}\sum_{i=1}^{n} x_i^* u_i + \left(\text{p lim}\frac{1}{n}\sum_{i=1}^{n} x_i^* e_i\right)\beta + \text{p lim}\frac{1}{n}\sum_{i=1}^{n} u_i e_i}{\text{p lim}\frac{1}{n}\sum_{i=1}^{n} x_i^{*2} + 2\text{p lim}\frac{1}{n}\sum_{i=1}^{n} x_i^* e_i + \text{p lim}\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

$$= \frac{\text{E}(x_i^{*2})\beta + \cancel{\text{E}(x_i^* u_i)}^{0} + \cancel{\text{E}(x_i^* e_i)}^{0}\beta + \cancel{\text{E}(u_i e_i)}^{0}}{\text{E}(x_i^{*2}) + 2\cancel{\text{E}(x_i^* e_i)}_{0} + \text{E}(e_i^2)} = \frac{\text{E}(x_i^{*2})\beta + 0 + 0 + 0}{\text{E}(x_i^{*2}) + 0 + \text{E}(e_i^2)}$$

$$= \underbrace{\left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2}\right)}_{<1}\beta = \frac{\beta}{1 + (\sigma_e^2/\sigma_{x^*}^2)} \neq \beta \qquad \text{if } \sigma_e^2 > 0$$

11

$$\underset{n\to\infty}{\text{p lim}}\ \widehat{\beta}_{\text{OLS}} = \frac{\beta}{1 + (\sigma_e^2/\sigma_{x^*}^2)} < \beta \qquad \text{for } \beta > 0 \text{ and } \sigma_e^2 > 0$$

$$\underset{n\to\infty}{\text{p lim}}\ \widehat{\beta}_{\text{OLS}} = \frac{\beta}{1 + (\sigma_e^2/\sigma_{x^*}^2)} > \beta \qquad \text{for } \beta < 0 \text{ and } \sigma_e^2 > 0$$

Since x_i is demeaned SLR

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^Ty^* = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}$$

$$\underset{n\to\infty}{\text{plim}}\ \hat{\beta}_{OLS} = \frac{\mathbb{E}[x_i y_i]}{\mathbb{E}[x_i^2]} = \frac{\mathbb{E}[(x_i^*+e_i)y_i]}{\mathbb{E}[(x_i^*+e_i)^2]}$$

$$= \frac{\mathbb{E}[(x_i^*+e_i)(x_i^*\beta+u_i)]}{\mathbb{E}[x_i^{*2} + 2x_i^*e_i + e_i^2]}$$

$$= \frac{\mathbb{E}[x_i^{*2}\beta] + \cancel{\mathbb{E}[x_i^* u_i]}^{0} + \cancel{\mathbb{E}[e_i x_i^*\beta]}^{0} + \cancel{\mathbb{E}[e_i u_i]}^{0}}{\mathbb{E}[x_i^{*2}] + 2\cancel{\mathbb{E}[x_i^* e_i]}_{0} + \mathbb{E}[e_i^2]}$$

$$= \frac{\beta\, \overbrace{\mathbb{E}[x_i^{*2}]}^{\sigma_{x^*}^2}}{\underbrace{\mathbb{E}[x_i^{*2}]}_{\sigma_{x^*}^2} + \underbrace{\mathbb{E}[e_i^2]}_{=\sigma_e^2 > 0}}$$

$$= \frac{\beta}{1 + \frac{\sigma_e^2}{\sigma_{x^*}^2}} \neq \beta \quad \text{if } \sigma_e^2 \neq 0$$

$$\begin{cases} \underset{n\to\infty}{\text{plim}}\ \hat{\beta}_{OLS} < \beta \quad \text{if } \beta > 0 \text{ and } \sigma_e^2 \neq 0 \\ \underset{n\to\infty}{\text{plim}}\ \hat{\beta}_{OLS} > \beta \quad \text{if } \beta < 0 \text{ and } \sigma_e^2 \neq 0 \end{cases}$$

- MLR
  - The OLS estimator of the coefficient on the variable with ME will have attenuation bias
  - The OLS estimator of other coefficients will also be biased, but with unknown directions

Omitted Variables

Omitted Variable in SLR

- True DGP:

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + u_i$$

  with $\mathbb{E}[x_{1i}u_i] = \mathbb{E}[x_{2i}u_i] = 0$ and $\mathbb{E}[u_i] = \mathbb{E}[x_{1i}] = \mathbb{E}[x_{2i}] = 0$ for simplicity
- We omit x_{2i} and estimate:

$$y_i = x_{1i}\beta_1 + (x_{2i}\beta_2 + u_i)$$

- Result:

$$\text{plim}_{n\to\infty}\hat{\beta}_1 = \beta_1 + \beta_2\rho_{x_1,x_2}$$

- Proof:

- The OLS estimator of β_1 is

  $$\widehat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$

  Substituting for $y = X_1\beta_1 + X_2\beta_2 + u$ from the true model

  $$\begin{aligned}\widehat{\beta}_1 &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + u) \\ &= \beta_1 + \left[(X_1'X_1)^{-1}X_1'X_2\right]\beta_2 + (X_1'X_1)^{-1}X_1'u \\ &= \beta_1 + \widehat{\delta}\beta_2 + (X_1'X_1)^{-1}X_1'u\end{aligned}$$

  where $\widehat{\delta} = (X_1'X_1)^{-1}X_1'X_2$ is the OLS estimator of ...

  21

  ...the coefficient δ in a linear projection of the omitted variable x_{2i} on the included variable x_{1i}, i.e.

  $$x_{2i} = x_{1i}\delta + e_i$$

  Taking probability limits, and using $\mathrm{E}(x_{1i}u_i) = 0$, we obtain

  $$\operatorname*{p\,lim}_{n\to\infty}\widehat{\beta}_1 = \beta_1 + (\operatorname*{p\,lim}_{n\to\infty}\widehat{\delta})\beta_2 = \beta_1 + \delta\beta_2$$

Omitted Variables in MLR #flashcard

- **Single Omitted Variable in MLR**
  - Model:

    $$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_{K-1}x_{K-1,i} + (\beta_K x_{Ki} + u_i)$$

    where $\mathbb{E}[x_{ki}u_i] = 0$ for $k = 1, \ldots, K$
  - Result:

    $$\text{plim}_{n\to\infty}\hat{\beta}_k = \beta_k + \beta_K \rho_{x_k, x_K | x_{i\neq K}}$$

    where $\rho_{x_k, x_K | x_{i\neq K}}$ is the partial correlation between x_k, x_K, which can be obtained as δ_k in the linear projection:

    $$x_{Ki} = \delta_1 + \delta_2 x_{2i} + \cdots + \delta_{K-1}x_{K-1,i} + v_i$$
  - Proof:

Single omitted variable

1 var omitted

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_{K-1} x_{K-1,i} + (\beta_K x_{Ki} + u_i) \quad (1)$$

where $\mathrm{E}(x_{ki} u_i) = 0$ for $k = 1, \ldots, K$ (and $x_{1i} = 1$ for all $i = 1, \ldots, n$)

[Relation to general model: $K_1 = K - 1$, $K_2 = 1$]

Linear projection of omitted x_{Ki} on all the included variables

$$x_{Ki} = \delta_1 + \delta_2 x_{2i} + \ldots + \delta_{K-1} x_{K-1,i} + v_i$$

s.t. $\mathrm{E}(x_{ki} v_i) = 0$ for $k = 1, \ldots, K-1$ (by definition of linear projection)

$$x_{Ki} = \delta_1 + \delta_2 x_{2i} + \ldots + \delta_{K-1} x_{K-1,i} + v_i$$

$$\beta_K x_{Ki} = \beta_K \delta_1 + \beta_K \delta_2 x_{2i} + \cdots + \beta_K \delta_{K-1} x_{K-1,i} + \beta_K v_i$$

Substitute this into (1)

Multiply x_{Ki} by β_K and subsititute:

$$y_i = (\beta_1 + \beta_K \delta_1) + (\beta_2 + \beta_K \delta_2) x_{2i} + \ldots + (\beta_{K-1} + \beta_K \delta_{K-1}) x_{K-1,i} + (u_i + \beta_K v_i)$$

29

$$y_i = (\beta_1 + \beta_K \delta_1) + (\beta_2 + \beta_K \delta_2) x_{2i} + \ldots + (\beta_{K-1} + \beta_K \delta_{K-1}) x_{K-1,i} + (u_i + \beta_K v_i)$$

Now since $\mathrm{E}[x_{ki}(u_i + \beta_K v_i)] = 0$ for $k = 1, \ldots, K-1$, we have

$\rho_{x_k, x_K | x_{j \neq k, K}}$

$$\operatorname{p}\lim_{n \to \infty} \widehat{\beta}_k = \beta_k + \beta_K \delta_k \qquad \text{for } k = 1, \ldots, K-1$$

$$\text{(Or equivalently } \operatorname{p}\lim_{n \to \infty} \widehat{\beta}_k = \beta_k + (\operatorname{p}\lim_{n \to \infty} \widehat{\delta}_k)\beta_K)$$

- **Multiple Omitted Variable in MLR**
  - Model:

$$y = X_1 \beta_1 + (X_2 \beta_2 + u)$$

  - We can show that:

$$\operatorname{plim}_{n \to \infty} \hat{\beta}_1 = \beta_1 + \left(\operatorname{plim}_{n \to \infty} (X_1^T X_1)^{-1} X_1^T X_2\right) \beta_2$$

  - OLS estimators will be biased and inconsistent (unless all omitted variables are orthogonal to all included variables) but the direction is hard to predict

Sinultaneity Bias #flashcard

- The dependent variable and at least one of the explanatory variables are chosen jointly as part of the same decision problem.

3 2 Instrument Variables

THE FOLLOWING IS NOT A COMPREHENSIVE SUMMARY FOR THE IV PART!

Formulas for 2SLS Estimator #flashcard

- $$\begin{aligned} \hat{\beta}_{2SLS} &= \left(\hat{X}^T \hat{X}\right)^{-1} \hat{X}^T y \\ &= \left(X^T \underbrace{Z(Z^T Z)^{-1} Z^T}_{P_z} X\right)^{-1} X^T \underbrace{Z(Z^T Z)^{-1} Z^T}_{P_z} y \\ &= \left(\hat{X}^T X\right)^{-1} \hat{X}^T y \\ &= (Z^T X)^{-1} Z^T y \qquad \text{(only in the just-identified case)} \end{aligned}$$

2SLS as Control Function #flashcard

1. Preform 1st stare projection
2. Plug in 1st-stage residuals as additional variables in the main regression

2SLS as Indirect Least Square #flashcard

- When *just-identified*, 2SLS coincides with indirect least squares
  - Example: we have an IV z_i for the endogenous regressor x_i
  1. Run a first stage projection

$$x_i = z_i\pi + r_i$$

  2. Run a reduced-form regression:

$$y_i = z_i d + u_i$$

  3. Divide reduced-form coefficient by 1st-stage coefficient:

$$\hat{\beta}_{\text{indirect square}} = \frac{\hat{d}}{\hat{\pi}}$$

Consistency of 2SLS #flashcard

- y of To establish conditions under which $\widehat{\beta}_{2SLS}$ is a consistent estimator of β, we express $\widehat{\beta}_{2SLS}$ in the form

$$\begin{aligned}\widehat{\beta}_{2SLS} &= (\widehat{X}'X)^{-1}\widehat{X}'y \quad \text{(3rd Expression)}\\ &= (\widehat{X}'X)^{-1}\widehat{X}'(X\beta + u)\\ &= (\widehat{X}'X)^{-1}\widehat{X}'X\beta + (\widehat{X}'X)^{-1}\widehat{X}'u\\ &= \beta + \left(\frac{\widehat{X}'X}{n}\right)^{-1}\left(\frac{\widehat{X}'u}{n}\right)\end{aligned}$$

Taking probability limits

$$\operatorname*{p\,lim}_{n\to\infty}\widehat{\beta}_{2SLS} = \beta + \operatorname*{p\,lim}_{n\to\infty}\left(\frac{\widehat{X}'X}{n}\right)^{-1}\operatorname*{p\,lim}_{n\to\infty}\left(\frac{\widehat{X}'u}{n}\right)$$

25

We assume the data on ① (y_i, x_i, z_i') are independent and identically distributed, with ② Validity $E(z_iu_i) = 0$ and ③ Information $E(z_ix_i) \neq 0 \leftrightarrow \pi \neq 0$

From the Law of Large Numbers, the vector of sample means

$$\frac{1}{n}\sum_{i=1}^{n} z_iu_i \xrightarrow{P} E(z_iu_i) = 0$$

We can also write $\frac{1}{n}\sum_{i=1}^{n} z_iu_i = \frac{1}{n}(Z'u)$, so we have

$$\left(\frac{Z'u}{n}\right) \xrightarrow{P} 0$$

$\widehat{\pi}$ is a consistent estimator of the coefficient vector π in the first stage linear projection, so we also have $\widehat{\pi} \xrightarrow{P} \pi \neq 0$

$$\begin{aligned}\hat{\beta}_{2SLS} &= (\hat{X}^TX)^{-1}\hat{X}^Ty\\ &= (\hat{X}^TX)^{-1}\hat{X}^T(X\beta + u)\\ &= (\hat{X}^TX)^{-1}\hat{X}^TX\beta + (\hat{X}^TX)^{-1}\hat{X}^Tu\\ &= \beta + (\hat{X}^TX)^{-1}\hat{X}^Tu\\ &= \beta + \left(\frac{\hat{X}^TX}{n}\right)^{-1}\left(\frac{\hat{X}^Tu}{n}\right)\end{aligned}$$

$$\operatorname*{plim}_{n\to\infty}\hat{\beta}_{2SLS} = \beta + \underbrace{\operatorname*{plim}_{n\to\infty}\left(\frac{\hat{X}^TX}{n}\right)^{-1}}\,\underbrace{\operatorname*{plim}_{n\to\infty}\left(\frac{\hat{X}^Tu}{n}\right)}$$

(show this is 0)

$$\frac{\hat{X}^Tu}{n} = \frac{(Z\hat{\pi})^Tu}{n} = \frac{\hat{\pi}^TZ^Tu}{n} = \hat{\pi}^T\frac{Z^Tu}{n}$$

$$\operatorname*{plim}_{n\to\infty}\left(\frac{\hat{X}^Tu}{n}\right) = \operatorname*{plim}_{n\to\infty}\left(\hat{\pi}^T\frac{Z^Tu}{n}\right) = \operatorname*{plim}_{n\to\infty}\hat{\pi}^T\operatorname*{plim}_{n\to\infty}\left(\frac{Z^Tu}{n}\right) = \pi^T\underbrace{E[Z_iu_i]}_{=0\text{ by Validity Assumption}} = \pi^T\times 0 = 0$$

(show this is finite)

$$\frac{\hat{X}^TX}{n} = \frac{(Z\hat{\pi})^TX}{n} = \frac{\hat{\pi}^TZ^TX}{n} = \hat{\pi}^T\left(\frac{Z^TX}{n}\right)$$

$$\operatorname*{plim}_{n\to\infty}\left(\frac{\hat{X}^TX}{n}\right) = \pi^T\operatorname*{plim}_{n\to\infty}\left(\frac{Z^TX}{n}\right) = \pi^T\underbrace{E[Z_iX_i]}_{\neq 0\text{ by Information Assumption}} = \pi^TM_{ZX} \neq 0$$

$$\Rightarrow \operatorname*{plim}_{n\to\infty}\hat{\beta}_{2SLS} = \beta + (\pi^TM_{ZX})^{-1}\times 0 = \beta \quad \square$$

- Since $\widehat{X} = Z\widehat{\pi}$, we can express

$$\left(\frac{\widehat{X}'u}{n}\right) = \left(\frac{(Z\widehat{\pi})'u}{n}\right) = \left(\frac{(\widehat{\pi}'Z')u}{n}\right) = \widehat{\pi}'\left(\frac{Z'u}{n}\right)$$

Then using Slutsky's theorem, we have

$$\operatorname*{p\,lim}_{n\to\infty}\left(\frac{\widehat{X}'u}{n}\right) = \pi'0 = 0$$

Similarly, from the Law of Large Numbers, the vector of sample means

$$\frac{1}{n}\sum_{i=1}^{n} z_ix_i \xrightarrow{P} E(z_ix_i) = M_{ZX} \neq 0$$

where $M_{ZX} = E(z_ix_i)$ is an $L \times 1$ column vector

27

We can also write $\frac{1}{n}\sum_{i=1}^{n} z_ix_i = \frac{1}{n}(Z'X)$, so we have

$$\left(\frac{Z'X}{n}\right) \xrightarrow{P} M_{ZX} \neq 0$$

Since $\widehat{X} = Z\widehat{\pi}$, we can express

$$\left(\frac{\widehat{X}'X}{n}\right) = \left(\frac{(Z\widehat{\pi})'X}{n}\right) = \left(\frac{(\widehat{\pi}'Z')X}{n}\right) = \widehat{\pi}'\left(\frac{Z'X}{n}\right)$$

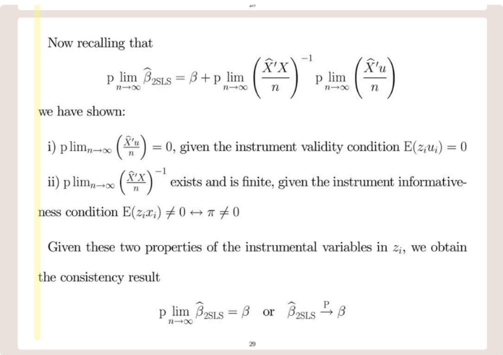
Then using Slutsky's theorem, we have

$$\operatorname*{p\,lim}_{n\to\infty}\left(\frac{\widehat{X}'X}{n}\right) = \pi'M_{ZX} \neq 0$$

and

$$\operatorname*{p\,lim}_{n\to\infty}\left(\frac{\widehat{X}'X}{n}\right)^{-1} = (\pi'M_{ZX})^{-1} \text{ is finite}$$

28

- 

Now recalling that

$$\operatorname{p}\lim_{n\to\infty} \widehat{\beta}_{2SLS} = \beta + \operatorname{p}\lim_{n\to\infty} \left(\frac{\widehat{X}'X}{n}\right)^{-1} \operatorname{p}\lim_{n\to\infty} \left(\frac{\widehat{X}'u}{n}\right)$$

we have shown:

i) $\operatorname{p}\lim_{n\to\infty} \left(\frac{\widehat{X}'u}{n}\right) = 0$, given the instrument validity condition $\mathrm{E}(z_i u_i) = 0$

ii) $\operatorname{p}\lim_{n\to\infty} \left(\frac{\widehat{X}'X}{n}\right)^{-1}$ exists and is finite, given the instrument informativeness condition $\mathrm{E}(z_i x_i) \neq 0 \leftrightarrow \pi \neq 0$

Given these two properties of the instrumental variables in z_i, we obtain the consistency result

$$\operatorname{p}\lim_{n\to\infty} \widehat{\beta}_{2SLS} = \beta \quad \text{or} \quad \widehat{\beta}_{2SLS} \xrightarrow{\mathrm{P}} \beta$$

29

2SLS as GMM #flashcard

- GMM tries to best match the sample analogue of population moment $\mathbb{E}\left[z_i u_i\right] = 0$
- Just identified case: $\hat{\beta}_{GMM} = \hat{\beta}_{2SLS}$
- Over-identified case: the GMM estimator minimises a weighted quadratic distance:

$$\begin{aligned} \hat{\beta}_{GMM} &= arg\min_{\beta} \left\{u^T Z W_n Z^T u\right\} \\ &= arg\min_{\beta} \left\{\left(\frac{1}{n}\sum_{i=1}^{n} u_i(\beta) z_i^T\right) W_n \left(\frac{1}{n}\sum_{i=1}^{n} u_i(\beta)^T z_i\right)\right\} \end{aligned}$$

- 2SLS uses a particular weight matrix $W_{2SLS} = (Z^T Z)^{-1}$, which is the most efficient one under homoskedascity

Inference and Var Estimation for 2SLS #flashcard

- Assumptions: Validity + Informative
- Large-sample distribution:

$$\hat{\beta}_{2SLS} \sim^a N\left(\beta, \frac{V}{n}\right)$$

- Under homoskedasticity ($\mathbb{E}\left[u_i^2 | z_i\right] = \sigma^2$), the consistent estimator for estimation variance is:

$$\widehat{Var}(\hat{\beta}_{2SLS}) = \frac{\hat{V}}{n} = \hat{\sigma}^2 \left(\hat{X}^T \hat{X}\right)^{-1}$$

where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2$
- Thus:

$$\hat{\beta}_{2SLS} \sim^a N\left(\beta, \hat{\sigma}^2 \left(\hat{X}^T \hat{X}\right)^{-1}\right)$$

- Under heteroskedasticity, there is a HR estimator:

$$\widehat{Var}_{HR}(\hat{\beta}_{2SLS}) = \frac{\hat{V}_{HR}}{n} = \left(\hat{X}^T \hat{X}\right)^{-1} \left(\sum_{i=1}^{n} \hat{u}_i^2 \hat{x}_i \hat{x}_i^T\right) \left(\hat{X}^T \hat{X}\right)^{-1}$$

- Thus:

$$\hat{\beta}_{2SLS} \sim^a N\left(\beta, \left(\hat{X}^T \hat{X}\right)^{-1} \left(\sum_{i=1}^{n} \hat{u}_i^2 \hat{x}_i \hat{x}_i^T\right) \left(\hat{X}^T \hat{X}\right)^{-1}\right)$$

- Note that $\hat{u}_i = y_i - {x_i}^T \hat{\beta}_{2SLS}$ (we use the true x_i not the predicted $\hat{x}_i$)

2SLS Procedures and Conditions for Multiple Endogenous Variables #flashcard

- Notations: L is the number of exogenous variables, K is the number of all variables in the equation of interest, $\underbrace{z_i^T}_{1\times L}$ is a row vector of all exogenous variables (IVs and exogenous regressors), $\underbrace{x_i^T}_{1\times K}$ is a row vector of all variables in the equation of interest
- 1st-stage Projection:

$$\underset{n\times K}{X} = \underset{n\times L}{Z}\,\underset{L\times K}{\Pi} + \underset{n\times K}{R}$$

note that if we have an intercept, we will alway include an equation $1 = 1$ and if some variables x_k are exogenous, we need to include $x_k = x_k$

- Conditions:
  - Validity: $\mathbb{E}\left[z_i u_i\right]=0$
  - Informative
    - Order condition (necessary but not sufficient): $L \geq K$
    - Rank condition (necessary and sufficient): the $L \times K$ matrix Π has full rank
- Then, calculate $\hat{X}$ and run the main regression

Testing IV Validity in Over-identifying Cases #flashcard

- Idea: when we have over-identification ($L>K$), we can examine whether the minimised value of the GMM criterion function is "small enough" to be consistent with our orthogonal assumption $\mathbb{E}\left[z_i u_i\right]=0$
- Conditional homoskedastic - **Sargan Test**
  - Sargan/J-test statistic:

$$J=\frac{1}{\hat{\sigma}^2} \hat{u}^T Z\left(Z^T Z\right)^{-1} Z^T \hat{u}$$

  where $\hat{u}_i=y_i-x_i^T \hat{\beta}_{2SLS}$ and $\hat{\sigma}^2=\frac{1}{n} \sum_{i=1}^n \hat{u}_i^2$
  - Distribution under $H_0: \mathbb{E}\left[z_i u_i\right]=0$:

$$J \sim^a \chi_{L-K}^2$$

- Conditional heteroskedastic - **Hansen Test**
  - No detail description

Testing Endogeneity #flashcard

- We assume the IVs are valid and informative
- Idea: $\hat{\beta}_{OLS}$ and $\hat{\beta}_{2SLS}$ should be similar if the variable is exogeous
- Conditional homoskedastic - **Hausman Test**
  - Assumption:
    - Valid and informative IVs
    - Conditional homoskedasticity:

$$\mathbb{E}\left[u_i^2 \mid z_i\right]=\sigma^2$$

    - The $K \times K$ matrix $\left(\widehat{Var}\left[\hat{\beta}_{2SLS}\right]-\widehat{Var}\left[\hat{\beta}_{OLS}\right]\right)$ is non-singular
      - If this is singular, we can use the Moore-Penrose pseudo-inverse with rank R, and the distribution will be $h \sim^a \chi_R^2$
  - $H_0: \mathbb{E}\left[x_i u_i\right]=0, H_1: \mathbb{E}\left[x_i u_i\right] \neq 0$
  - Test statistic:

$$h=\left(\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}\right)^T\left(\widehat{Var}\left[\hat{\beta}_{2SLS}\right]-\widehat{Var}\left[\hat{\beta}_{OLS}\right]\right)^{-1}\left(\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}\right)$$

  - Distribution under H_0:

$$h \sim^a \chi_K^2$$

  - If we are only interested in a *sub-vector* of β with K_1 variables, we simply repeat the above with our sub-vector, and the distribution will be $\chi_{K_1}^2$
  - If we are only interested in *one parameter* β_k, then the test simplifies to:

$$h=\frac{\left(\hat{\beta}_{k,2SLS}-\hat{\beta}_{k,OLS}\right)^2}{\widehat{Var}\left[\hat{\beta}_{k,2SLS}\right]-\widehat{Var}\left[\hat{\beta}_{k,OLS}\right]} \sim^a \chi_1^2$$

- Alternative method (can deal with heteroskedasticity easily): **Control Function Test**
  - Perform 2SLS estimation using the control function method
  - If there is only 1 endogenous variable of interest: use a t-test to test whether the coefficient on the 1st-stage residual is 0 in the 2nd-stage regression
  - If there are more than 1 endogenous variables of interest: use a Wald test to test whether all coefficients on the 1st-stage residual are jointly 0 in the 2nd-stage regression
  - We can easily deal with heteroskedasticity using HR estimtor of variance, but testing on a sub-vector of β will be hard due to "generated regressors" problem.

Finite-Sample Problems #flashcard

- Overfitting: $\hat{\beta}_{2SLS} \rightarrow \hat{\beta}_{OLS}$ as $L \rightarrow n$
  - A simple way to investigate: calculate a sequance of 2SLS estimates based on smaller and smaller subsets of the original IVs, and check whether there's systematic tendency for $\hat{\beta}_{2SLS}$ to move away from $\hat{\beta}_{OLS}$

Weak Instruments #flashcard

- Finite sample bias + Large inconsistency

Tests for Weak Instruments #flashcard

- Run a Wald Test for the 1st stage: $H_0 : \delta_0 m = \delta_1 = \cdots = \delta_M = 0$
- Test statistics $\sim F(M, n-L)$ and we typically require it to be greater than 10.

Weak IV Robust Inference: Anderson-Rubin Test #flashcard

- The Anderson-Rubin test is a robust test for the significance of endogenous regressors in IV models, and unlike the usual Wald or t-tests, it remains valid even when instruments are weak.
  - It tests the null hypothesis:

$$H_0 : \beta = \beta_0$$

  where β is the coefficient on the endogenous regressor.
- **Procedures**
  - We start with the model:

$$y = X\beta + \varepsilon$$

and instrument X using Z (with $\text{rank}(Z) = m$, number of instruments).
- Instead of relying on 2SLS estimates, the AR test does the following:
1. Move the hypothesized value to the left-hand side:

$$y - X\beta_0 = u$$

2. Test if the residual u is uncorrelated with the instruments Z:

$$H_0 : \mathbb{E}[Z^T u] = 0$$

3. Form the test statistic:

$$AR(\beta_0) = \frac{\widehat{u}^T P_Z \widehat{u}}{\widehat{\sigma}^2}$$

where:
- $P_Z = Z(Z^T Z)^{-1} Z^T$ is the projection matrix onto the instrument space.
- $\hat{\sigma}^2$ is an estimator of the error variance (often from a reduced form).
4. Distribution under the null:

$$AR(\beta_0) \sim \chi^2_m$$

where $m =$ number of instruments.

3 4 2 Local Average Treatment Effects

3 5 GMM

GMM Estimator #flashcard

- Setup:

$$y_i - x_i^T \beta = u_i(\beta),\ \mathbb{E}[u_i] = 0, \mathbb{E}[z_i u_i] = 0$$

- GMM Estimator:

$$\hat{\beta}_{GMM} = arg \min_{\beta} \hat{u}^T Z W_n Z^T \hat{u}$$

- Expressions and Weak Consistency:
  - Assumptions:
  - $y_i = x_i^T \beta + u_i$
  - (x_i, y_i, z_i) are idd
  - $Rank(\mathbb{E}\left[z_i x_i^T\right]) = K$
  - $W_n \rightarrow^p W$ is a symmetric and psd $L \times L$ matrix
  - Then:

$$\begin{aligned}\hat{\beta}_{GMM} &= \left((X^T Z) W_n \, (Z^T X) \right)^{-1} (X^T Z) W_n \, (Z^T y) \\ &\rightarrow^p \beta\end{aligned}$$

  - and:

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \rightarrow^D N(0, V)$$