Advanced Econometrics

Lecture 9: Instrumental Variables (Hansen Chapter 11)

Instructor: Ma, Jun

Renmin University of China

December 13, 2021

Introduction

- Endogeneity in the linear model:

$$\begin{aligned} Y_i &= \boldsymbol{X}_i'\boldsymbol{\beta} + e_i \\ \mathbb{E}\left(\boldsymbol{X}_i e_i\right) &\neq \mathbf{0}. \end{aligned}$$

- Note that the above model is not the linear projection model, since otherwise, if $\boldsymbol{\beta}^* = \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i Y_i\right)$, and the linear projection model is

$$\begin{aligned} Y_i &= \boldsymbol{X}_i'\boldsymbol{\beta}^* + e_i^* \\ \mathbb{E}\left(\boldsymbol{X}_i e_i^*\right) &= \mathbf{0}. \end{aligned}$$

- We always assume that $\mathbb{E}\left(e_i\right) = 0$ and the first coordinate of $\boldsymbol{X}_i$ is 1 so that its coefficient is the intercept. Under this assumption, $\mathbb{E}\left(\boldsymbol{X}_i e_i\right) \neq \mathbf{0}$ if and only if e_i is correlated with one of the regressors.

- Under endogeneity, the projection coefficients $\boldsymbol{\beta}^*$ does not equal the structural parameter $\boldsymbol{\beta}$:

$$\begin{aligned}
\boldsymbol{\beta}^* &= \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i Y_i\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i \left(\boldsymbol{X}_i'\boldsymbol{\beta} + e_i\right)\right) \\
&= \boldsymbol{\beta} + \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i e_i\right) \\
&\neq \boldsymbol{\beta}.
\end{aligned}$$

- Endogeneity implies that the LS estimator is inconsistent for the structural parameter $\boldsymbol{\beta}$. The LS estimator is consistent for the projection coefficient $\boldsymbol{\beta}^*$:

$$\widehat{\boldsymbol{\beta}} \to_p \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{X}_i Y_i\right) = \boldsymbol{\beta}^* \neq \boldsymbol{\beta}.$$

The simple case of one regressor ($k = 1$)

- Consider

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + e_i, \\
\mathrm{E}[e_i] &= 0 \\
\mathrm{Cov}[X_i, e_i] &\neq 0.
\end{aligned}
$$

- An instrument is an variable Z_i which satisfies the following conditions:
  1. The IV is exogenous: $\mathrm{Cov}[Z_i, e_i] = 0$.
  2. The IV determines the endogenous regressor: $\mathrm{Cov}[Z_i, X_i] \neq 0$.
- When an IV variable satisfying those conditions is available, it allows us to estimate the effect of X on Y consistently.

Sources of endogeneity

There are several possible sources of endogeneity:

1. Omitted explanatory variables.
2. Simultaneity.
3. Errors in variables.

All result in regressors correlated with the errors.

Omitted explanatory variables

- Suppose that the true model is

$$\ln Wage_i = \beta_0 + \beta_1 Education_i + \beta_2 Ability_i + V_i,$$

  where V_i is uncorrelated with $Education$ and $Ability$.
- Since $Ability$ is unobservable, the econometrician regresses $\ln Wage$ against $Education$, and $\beta_2 Ability$ goes into the error part:

$$\begin{aligned} \ln Wage_i &= \beta_0 + \beta_1 Education_i + U_i, \\ U_i &= \beta_2 Ability_i + V_i. \end{aligned}$$

- $Education$ is correlated with $Ability$: we can expect that $\mathrm{Cov}\left(Education_i, Ability_i\right) > 0, \beta_2 > 0$, and therefore $\mathrm{Cov}\left(Education_i, U_i\right) > 0$.

Simultaneity

- Consider the following demand-supply system:

$$\begin{array}{ll} \text{Demand:} & Q^d = \beta_0^d + \beta_1^d P + U^d, \\ \text{Supply:} & Q^s = \beta_0^s + \beta_1^s P + U^s, \end{array}$$

  where: Q^d =quantity demanded, Q^s =quantity supplied, P=price.
- The quantity and price are determined simultaneously in the equilibrium:

$$Q^d = Q^s = Q.$$

- Note that Q^d and Q^s are not observed separately, we observe only the equilibrium values Q.

$$
\begin{aligned}
Q^d &= \beta_0^d + \beta_1^d P + U^d, \\
Q^s &= \beta_0^s + \beta_1^s P + U^s, \\
Q^d &= Q^s = Q.
\end{aligned}
$$

- Solving for P, we obtain

$$
0 = \left(\beta_0^d - \beta_0^s\right) + \left(\beta_1^d - \beta_1^s\right) P + \left(U^d - U^s\right),
$$

  or

$$
P = -\frac{\beta_0^d - \beta_0^s}{\beta_1^d - \beta_1^s} - \frac{U^d - U^s}{\beta_1^d - \beta_1^s}.
$$

- Thus,

$$
\operatorname{Cov}\left(P, U^d\right) \neq 0 \text{ and } \operatorname{Cov}\left(P, U^s\right) \neq 0.
$$

  The demand-supply equations cannot be estimated by OLS.

- Consider the following labour supply model for married women:

$$Hours_i = \beta_0 + \beta_1 Children_i + \text{Other Factors} + U_i,$$

  where $Hours$=hours of work, $Children$=number of children.
- It is reasonable to assume that women decide simultaneously how much time to devote to career and family.
- Thus, while we may be mainly interested in the effect of family size on labour supply, there is another equation:

$$Children_i = \gamma_0 + \gamma_1 Hours_i + \text{Other Factors} + V_i,$$

  and $Children$ and $Hours$ are determined simultaneously in an equilibrium.
- As a result, $\text{Cov}\left(Children_i, U_i\right) \neq 0$, and the effect of family size cannot be estimated by OLS.

Errors in variables

- Consider the following model:

$$Y_i = \beta_0 + \beta_1 X_i^* + V_i,$$

where X_i^* is the true regressor.

- Suppose that X_i^* is not directly observable. Instead, we observe X_i that measures X_i^* with an error ε_i:

$$X_i = X_i^* + \varepsilon_i.$$

- Since X_i^* is unobservable, the econometrician has to regress Y_i against X_i.

$$
\begin{aligned}
X_i &= X_i^* + \varepsilon_i, \\
Y_i &= \beta_0 + \beta_1 X_i^* + V_i.
\end{aligned}
$$

- The model for Y_i as a function of X_i can be written as

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 \left(X_i - \varepsilon_i \right) + V_i \\
&= \beta_0 + \beta_1 X_i + V_i - \beta_1 \varepsilon_i,
\end{aligned}
$$

or

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + e_i, \\
e_i &= V_i - \beta_1 \varepsilon_i.
\end{aligned}
$$

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + e_i, \\
e_i &= V_i - \beta_1 \varepsilon_i, \\
X_i &= X_i^* + \varepsilon_i.
\end{aligned}
$$

- We can assume that

$$\mathrm{Cov}\left[X_i^*, V_i\right] = \mathrm{Cov}\left[X_i^*, \varepsilon_i\right] = \mathrm{Cov}\left[\varepsilon_i, V_i\right] = 0.$$

- However,

$$
\begin{aligned}
\mathrm{Cov}\left[X_i, e_i\right] &= \mathrm{Cov}\left[X_i^* + \varepsilon_i, V_i - \beta_1 \varepsilon_i\right] \\
&= \mathrm{Cov}\left[X_i^*, V_i\right] - \beta_1 \mathrm{Cov}\left[X_i^*, \varepsilon_i\right] \\
&\quad + \mathrm{Cov}\left[\varepsilon_i, V_i\right] - \beta_1 \mathrm{Cov}\left[\varepsilon_i, \varepsilon_i\right]
\end{aligned}
$$

- Thus, X_i is enodgenous and β_1 cannot be estimated by OLS.

- Theoretically, the causal effect can be estimated from controlled experiments:
  - To estimate the return to education, select a random sample of children, randomly assign how many years of education they should have, and measure their income several years after the graduation.
  - To estimate the effect of family size on labor supply, select a random sample of parents and randomly assign how many children they should have, and measure their labor market outcomes.
  - Such an approach is infeasible due to a high cost and/or ethical reasons.
- Natural experiments: Use the random variation in the variable of interest to estimate the causal effect.

Example: Compulsory schooling laws and return to education

- Angrist and Krueger, 1991, *QJE*, suggested using school start age policy to estimate β_1 in $\ln Wage_i = \beta_0 + \beta_1 Education_i + \beta_2 Ability_i + V_i$.
- We need to find an IV variable Z such that $\mathrm{Cov}\left(Ability_i, Z_i\right) = 0$ and $\mathrm{Cov}\left(Education_i, Z_i\right) \neq 0$.
- They argue that due to compulsory schooling laws, the season of birth variable satisfies the IV conditions:
  - A child has to attend the school until he reaches a certain drop-out age.
  - Students born in the first quarter of the year, reach the legal drop-out age before their classmates who were born later in the year.
  - The quarter of birth dummy variable is correlated with education.
  - The quarter of birth is uncorrelated with ability.

Example: Sibling-sex composition and labor supply

- Angrist and Evans, 1998, *AER*, argue that the parents' preferences for a mixed sibling-sex composition can be used to estimate β_1 in $Hours_i = \beta_0 + \beta_1 Children_i + \ldots + U_i$.
- We need to find an IV Z such that $\text{Cov}\left[U_i, Z_i\right] = 0$ and $\text{Cov}\left(Children_i, Z_i\right) \neq 0$.
- Consider a dummy variable that takes on the value one if the sex of the second child matches the sex of the first child.
  - If the parents prefer a mixed sibling-sex composition, they are more likely to have another child if their first two children are of the same sex.
  - The same-sex dummy is correlated with the number of children.
  - Since sex mix is randomly determined, the same sex dummy is exogenous.

Instrumental Variables

- Partition:

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{1i} \\ \boldsymbol{X}_{2i} \end{pmatrix} \begin{array}{l} k_1 \\ k_2 \end{array}$$

  and

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \begin{array}{l} k_1 \\ k_2 \end{array} .$$

- So the model is:

$$\begin{aligned} Y_i &= \boldsymbol{X}_i' \boldsymbol{\beta} + e_i \\ &= \boldsymbol{X}_{1i}' \boldsymbol{\beta}_1 + \boldsymbol{X}_{2i}' \boldsymbol{\beta}_2 + e_i . \end{aligned}$$

  In matrix notation:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \\ &= \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e} . \end{aligned}$$

- Assume

$$\mathbb{E}(\boldsymbol{X}_{1i}e_i) = \boldsymbol{0}$$
$$\mathbb{E}(\boldsymbol{X}_{2i}e_i) \neq \boldsymbol{0}$$

Definition

The $l \times 1$ random vector $\boldsymbol{Z}_i$ is an instrumental variable if

$$\begin{aligned} \mathbb{E}(\boldsymbol{Z}_i e_i) &= \boldsymbol{0} \\ \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{Z}_i'\right) &> 0 \\ \operatorname{rank}\left(\mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right)\right) &= k \end{aligned}$$

- $\boldsymbol{X}_{1i}$ satisfies $\mathbb{E}\left(\boldsymbol{X}_{1i}e_i\right) = \mathbf{0}$. So it should be included as instrumental variables.

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{Z}_{1i} \\ \boldsymbol{Z}_{2i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{1i} \\ \boldsymbol{Z}_{2i} \end{pmatrix} \begin{array}{c} k_1 \\ l_2 \end{array}$$

- We say the model is just-identified if $\ell = k$ ($\ell_2 = k_2$) and over-identified if $\ell > k$ ($\ell_2 > k_2$).

Instrumental Variables Estimator

- The assumption that $\boldsymbol{Z}_i$ is an IV implies

$$\mathbb{E}\left(\boldsymbol{Z}_i e_i\right) = \mathbf{0}$$
$$\mathbb{E}\left(\boldsymbol{Z}_i\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta}\right)\right) = \mathbf{0}$$
$$\mathbb{E}\left(\boldsymbol{Z}_i Y_i\right) - \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right)\boldsymbol{\beta} = \mathbf{0}.$$

- If $\ell = k$, solve for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \left(\mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{Z}_i Y_i\right).$$

- The IV estimator:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathrm{iv}} &= \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i \boldsymbol{X}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i Y_i\right) \\
&= \left(\sum_{i=1}^{n} \mathbf{Z}_i \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{Z}_i Y_i\right) \\
&= \left(\mathbf{Z}'\boldsymbol{X}\right)^{-1} \left(\mathbf{Z}'\boldsymbol{Y}\right).
\end{aligned}
$$

- The residual satisfies:

$$
\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{iv}}
$$

$$
\boldsymbol{Z}'\widehat{\boldsymbol{e}} = \boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\left(\mathbf{Z}'\boldsymbol{X}\right)^{-1}\left(\mathbf{Z}'\boldsymbol{Y}\right) = \mathbf{0}.
$$

Two-Stage Least Squares

- We denote $\widehat{\boldsymbol{\Gamma}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{X})$.

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\text{2sls}} &= \left(\widehat{\boldsymbol{\Gamma}}'\boldsymbol{Z}'\boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}\right)^{-1}\left(\widehat{\boldsymbol{\Gamma}}'\boldsymbol{Z}'\boldsymbol{Y}\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1} \\
&\quad \cdot \boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}.
\end{aligned}
$$

- When $k = \ell$, the 2SLS simplifies to IV:

$$
\begin{aligned}
\left(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1} &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\left((\boldsymbol{Z}'\boldsymbol{Z})^{-1}\right)^{-1}(\boldsymbol{X}'\boldsymbol{Z})^{-1} \\
&= (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{X}'\boldsymbol{Z})^{-1}
\end{aligned}
$$

- So

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\text{2sls}} &= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)\left(\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \widehat{\boldsymbol{\beta}}_{\text{iv}}.
\end{aligned}
$$

- Define the projection matrix:

$$
\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'.
$$

- We can write

$$
\widehat{\boldsymbol{\beta}}_{\text{2sls}} = \left(\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{Y}.
$$

- And the fitted values:

$$\widehat{\boldsymbol{X}} = \boldsymbol{P_Z X} = \mathbf{Z}\widehat{\boldsymbol{\Gamma}}$$

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{2sls}} &= \left(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{P_Z}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{Y} \\ &= \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{Y}. \end{aligned}$$

- First regress $\boldsymbol{X}$ on $\boldsymbol{Z}$. Obtain the LS coefficients $\widehat{\boldsymbol{\Gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\boldsymbol{X})$ and the fitted values $\widehat{\boldsymbol{X}} = \boldsymbol{P_Z X} = \mathbf{Z}\widehat{\boldsymbol{\Gamma}}$.
- Second regress $\boldsymbol{Y}$ on $\widehat{\boldsymbol{X}}$. Get $\widehat{\boldsymbol{\beta}}_{\text{2sls}} = \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{Y}$.

- Recall $\boldsymbol{X} = [\boldsymbol{X}_1\ \boldsymbol{X}_2]$ and $\mathbf{Z} = [\boldsymbol{X}_1\ \mathbf{Z}_2]$. Note $\widehat{\boldsymbol{X}}_1 = \boldsymbol{P}_{\mathbf{Z}}\boldsymbol{X}_1 = \boldsymbol{X}_1$. Then
$$\widehat{\boldsymbol{X}} = \left[\widehat{\boldsymbol{X}}_1, \widehat{\boldsymbol{X}}_2\right] = \left[\boldsymbol{X}_1, \widehat{\boldsymbol{X}}_2\right].$$
- The 2SLS residuals:
$$\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\text{2sls}}.$$
- When the model is overidentified, $\mathbf{Z}'\widehat{\boldsymbol{e}} \neq \mathbf{0}$ but
$$\begin{aligned}
\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{e}} &= \widehat{\boldsymbol{\Gamma}}'\mathbf{Z}'\widehat{\boldsymbol{e}} \\
&= \boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\widehat{\boldsymbol{e}} \\
&= \boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{Y} - \boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\text{2sls}} \\
&= \mathbf{0}.
\end{aligned}$$

Consistency of 2SLS

Assumption

1. The observations $(Y_i, \boldsymbol{X_i}, \mathbf{Z_i})$*,* $i = 1, \ldots, n$*, are independent and identically distributed.*

2. $\mathbb{E}\left(Y^2\right) < \infty$.

3. $\mathbb{E} \parallel \boldsymbol{X} \parallel^2 < \infty$.

4. $\mathbb{E} \parallel \boldsymbol{Z} \parallel^2 < \infty$.

5. $\mathbb{E}\left(\boldsymbol{Z}'\right)$ *is positive definite.*

6. $\mathbb{E}\left(\mathbf{Z}\boldsymbol{X}'\right)$ *has full rank k.*

7. $\mathbb{E}\left(\mathbf{Z}\boldsymbol{e}\right) = 0$.

- Proof of consistency:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{2sls}} &= \left(\boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\boldsymbol{Z}'\left(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e}\right) \\
&= \boldsymbol{\beta}+\left(\boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{e}.
\end{aligned}
$$

- Then

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{2sls}}-\boldsymbol{\beta} &= \left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1} \\
&\quad\cdot\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{e}\right).
\end{aligned}
$$

- Then,

$$\hat{\boldsymbol{\beta}}_{\text{2sls}} - \boldsymbol{\beta} \rightarrow_p \left(\boldsymbol{Q}_{XZ} \boldsymbol{Q}_{ZZ}^{-1} \boldsymbol{Q}_{ZX} \right)^{-1} \boldsymbol{Q}_{XZ} \boldsymbol{Q}_{ZZ}^{-1} \mathbb{E}\left(\mathbf{Z}_i e_i \right) = 0,$$

where

$$\begin{aligned} \boldsymbol{Q}_{XZ} &= \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{Z}_i' \right) \\ \boldsymbol{Q}_{ZZ} &= \mathbb{E}\left(\mathbf{Z}_i \boldsymbol{Z}_i' \right) \\ \boldsymbol{Q}_{ZX} &= \mathbb{E}\left(\mathbf{Z}_i \boldsymbol{X}_i' \right). \end{aligned}$$

Asymptotic Distribution of 2SLS

Assumption

1. $\mathbb{E}\left(Y^4\right) < \infty$.

2. $\mathbb{E} \parallel \boldsymbol{Z} \parallel^4 < \infty$.

- Write

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2\text{sls}} - \boldsymbol{\beta}) = \left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1} \cdot \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{e}\right).$$

- By CLT,

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{e} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{Z}_i e_i \to_d \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right),$$

where $\boldsymbol{\Omega} = \mathbb{E}\left(e_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'\right)$.

- Slutsky's theorem:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2\text{sls}} - \boldsymbol{\beta}) \rightarrow_d \left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\mathrm{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right) = \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right).$$

- We can verify:

$$\begin{aligned}\left(\mathbb{E}\left(e^4\right)\right)^{1/4} &= \left(\mathbb{E}\left((Y - \boldsymbol{X}'\boldsymbol{\beta})^4\right)\right)^{1/4} \\ &\leq \left(\mathbb{E}\left(Y^4\right)\right)^{1/4} + \parallel \boldsymbol{\beta} \parallel \left(\mathbb{E} \parallel \boldsymbol{X} \parallel^4\right)^{1/4} < \infty\end{aligned}$$

$$\mathbb{E} \parallel \boldsymbol{Z}e \parallel^2 \leq \left(\mathbb{E} \parallel \boldsymbol{Z} \parallel^4\right)^{1/2}\left(\mathbb{E}\left(e^4\right)\right)^{1/2} < \infty.$$

So the CLT and Slutsky's theorem do apply.

Theorem

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{2\text{sls}} - \boldsymbol{\beta}\right) \rightarrow_d \mathrm{N}\left(\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}}\right)$$

where

$$\mathbf{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}\left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right) \cdot \left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}$$

and

$$\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i' e_i^2\right).$$

- The asymptotic variance simplifies under a conditional homoskedasticity condition: $\mathbb{E}\left(e_i^2 | \boldsymbol{Z}_i\right) = \sigma^2$.
- $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}}^0 = \left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}\sigma^2.$

Covariance Matrix Estimation

- Estimator of the asymptotic variance matrix $\boldsymbol{V}_{\boldsymbol{\beta}}$:

$$\hat{\mathbf{V}}_{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{-1}\hat{\boldsymbol{Q}}_{\boldsymbol{ZX}}\right)^{-1}\left(\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{-1}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{-1}\hat{\boldsymbol{Q}}_{\boldsymbol{ZX}}\right) \cdot \left(\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{-1}\hat{\boldsymbol{Q}}_{\boldsymbol{ZX}}\right)^{-1}$$

where

$$\begin{aligned}
\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i' = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z} \\
\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i' = \frac{1}{n}\boldsymbol{X}'\boldsymbol{Z} \\
\hat{\boldsymbol{\Omega}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i'\hat{e}_i^2 \\
\hat{e}_i &= Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}_{\text{2sls}}.
\end{aligned}$$

- The homoskedastic variance matrix can be estimated by

$$
\begin{aligned}
\hat{\mathbf{V}}_{\boldsymbol{\beta}}^{0} &= \left(\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{\mathbf{-1}}\hat{\boldsymbol{Q}}_{\boldsymbol{ZX}}\right)^{\mathbf{-1}}\hat{\sigma}^{2} \\
\hat{\sigma}^{2} &= \frac{1}{n}\sum_{i=1}^{n}\hat{e}_{i}^{2}.
\end{aligned}
$$

Theorem

$$\hat{\mathbf{V}}_{\boldsymbol{\beta}}^{0} \rightarrow_{p} \mathbf{V}_{\boldsymbol{\beta}}^{0}$$

$$\hat{\mathbf{V}}_{\boldsymbol{\beta}} \rightarrow_{p} \mathbf{V}_{\boldsymbol{\beta}}.$$

- The covariance matrix estimator should be constructed using the correct residual formula: $\widehat{e}_i = Y_i - \boldsymbol{X}_i' \widehat{\boldsymbol{\beta}}_{\mathrm{2sls}}$.
- In the second stage, regress Y_i on $\widehat{\boldsymbol{X}}_i$, $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{\Gamma}}' \boldsymbol{Z}_i$.
- Residuals from the second stage: $Y_i = \widehat{\boldsymbol{X}}_i' \widehat{\boldsymbol{\beta}}_{\mathrm{2sls}} + \hat{v}_i$.
- The standard errors reported by STATA for the second-stage regression use the residual $\hat{v}_i$. The (homoskedastic) formula it uses is

$$\begin{aligned}
\hat{\mathbf{V}}_{\boldsymbol{\beta}} &= \left(\frac{1}{n} \widehat{\boldsymbol{X}}' \widehat{\boldsymbol{X}} \right)^{\mathbf{-1}} \hat{\sigma}_v^2 = \left(\hat{\boldsymbol{Q}}_{\boldsymbol{XZ}} \hat{\boldsymbol{Q}}_{\boldsymbol{ZZ}}^{\mathbf{-1}} \hat{\boldsymbol{Q}}_{\boldsymbol{ZX}} \right)^{-1} \hat{\sigma}_v^2 \\
\hat{\sigma}_v^2 &= \frac{1}{n} \sum_{i=1}^{n} \hat{v}_i^2 .
\end{aligned}$$

- However,

$$\begin{aligned}
\hat{v}_i &= Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_{\mathrm{2sls}} + \left(\boldsymbol{X}_i - \hat{\boldsymbol{X}}_i \right)' \hat{\boldsymbol{\beta}}_{\mathrm{2sls}} \\
&\neq \hat{e}_i .
\end{aligned}$$

Functions of Parameters

- Given $\boldsymbol{r} : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q$, the parameter of interest is $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$.
- A natural estimator is $\widehat{\boldsymbol{\theta}}_{\text{2sls}} = \boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}_{\text{2sls}}\right)$.

Theorem

r is continuous at $\boldsymbol{\beta}$, then $\hat{\boldsymbol{\theta}}_{\text{2sls}} \to_p \boldsymbol{\theta}$ as $n \to \infty$.

- Estimator of the asymptotic variance matrix:

$$\begin{aligned} \hat{\mathbf{V}}_{\theta} &= \hat{\boldsymbol{R}}' \hat{\mathbf{V}}_{\boldsymbol{\beta}} \hat{\boldsymbol{R}} \\ \hat{\boldsymbol{R}} &= \left. \frac{\partial}{\partial \boldsymbol{b}} r(\boldsymbol{b})' \right|_{\boldsymbol{b} = \hat{\boldsymbol{\beta}}_{\text{2sls}}} \end{aligned}$$

Theorem

If r is continuously differentiable at $\boldsymbol{\beta}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{2\text{sls}} - \boldsymbol{\theta}\right) \longrightarrow_d \mathrm{N}\left(\mathbf{0}, \mathbf{V}_\theta\right)$$

where

$$\begin{aligned} \mathbf{V}_\theta &= \boldsymbol{R}'\mathbf{V}_{\boldsymbol{\beta}}\boldsymbol{R} \\ \boldsymbol{R} &= \left.\frac{\partial}{\partial \boldsymbol{b}} r\left(\boldsymbol{b}\right)'\right|_{\boldsymbol{b}=\boldsymbol{\beta}} \end{aligned}$$

and $\hat{\mathbf{V}}_\theta \longrightarrow_p \mathbf{V}_\theta$.

Hypothesis Tests

- We are interested in testing

$$\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$$
$$\mathbb{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

- The Wald statistic:

$$W = n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \hat{\mathbf{V}}_{\hat{\theta}}^{-1}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

Theorem

$$W \longrightarrow_d \chi_q^2.$$

For c satisfying $\alpha = 1 - G_q(c)$,

$$\Pr(W > c \mid \mathbb{H}_0) \rightarrow \alpha$$

so the test "Reject $\mathbb{H}_0$ if $W > c$" has asymptotic size α.