Econometric Analysis of Cross Section and Panel Data

Lecture 8: Instrumental Variables

Zhian Hu

Central University of Finance and Economics

Fall 2024

This Lecture

- Hansen (2022): Chapter 12
- Instrumental Variable

Overview

- We say that there is endogeneity in the linear model

$$Y = X'\beta + e$$

  if β is the parameter of interest and

$$\mathbb{E}[Xe] \neq 0$$

- To distinguish from the regression and projection models, we will call it a structural equation and β a structural parameter.

Endogenous Regressors

- A regressor X is exogenous for β if $\mathbb{E}[Xe] = 0$.
- A regressor X is endogeneous for β if $\mathbb{E}[Xe] \neq 0$.
- In most applications only a subset of the regressors are treated as endogenous.
- Partition $X = (X_1, X_2)$ with dimensions (k_1, k_2) so that X_1 contains the exogenous regressors and X_2 contains the endogenous regressors.

$$Y = X_1'\beta_1 + X_2'\beta_2 + e$$

Endogenous Regressors

- An alternative notation is we let $Y_2 = X_2$ be the endogenous regressors and rename the dependent variable Y as Y_1. Then the structural equation is

$$Y_1 = X_1'\beta_1 + Y_2'\beta_2 + e$$

  - This is especially useful so that the notation clarifies which variables are endogenous and which exogenous.
- We also write $\vec{Y} = (Y_1, Y_2)$ as the set of endogenous variables.
- The assumptions regarding the regressors and regression error are

$$\mathbb{E}[X_1 e] = 0$$
$$\mathbb{E}[Y_2 e] \neq 0$$

Instruments

- **Definition:** The $\ell \times 1$ random vector Z is an instrumental variable for the structural equation if
$$\begin{aligned} \mathbb{E}[Ze] &= 0 \\ \mathbb{E}\left[ZZ'\right] &> 0 \\ \operatorname{rank}\left(\mathbb{E}\left[ZX'\right]\right) &= k \end{aligned}$$
- The first is that the instruments are uncorrelated with the regression error, which means they are **exogenous**.
- The second is a normalization which excludes linearly redundant instruments.
- The third is often called the **relevance condition** and is essential for the identification of the model.
    - A necessary condition is that $\ell \geq k$.

Instruments

- Notice that the regressors X_1 satisfy the first condition and thus should be included as instrumental variables. They are therefore a subset of the variables Z.
- Notationally we make the partition

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ Z_2 \end{pmatrix} \begin{array}{l} k_1 \\ \ell_2 \end{array}$$

- Here, $X_1 = Z_1$ are the **included exogenous variables** and Z_2 are the **excluded exogenous variables**.
- With this notation we can also write the structural equation as

$$Y_1 = Z_1'\beta_1 + Y_2'\beta_2 + e$$

- We say that the model is just-identified if $\ell = k$ and over-identified if $\ell > k$.

An IV Example: College Proximity

- David Card (1995) uses college proximity to instrument for year of schooling and estimates its economic return.
- He argues that a potential student lives close to a college this reduces the cost of attendance and thereby raises the likelihood that the student will attend college.
- College proximity does not directly affect a student's skills or abilities so should not have a direct effect on his or her market wage.
- These considerations suggest that college proximity can be used as an instrument for education in a wage regression.

An IV Example: College Proximity

	OLS	IV(a)	IV(b)	2SLS(a)	2SLS(b)	LIML
education	0.074	0.132	0.133	0.161	0.160	0.164
	(0.004)	(0.049)	(0.051)	(0.040)	(0.041)	(0.042)
experience	0.084	0.107	0.056	0.119	0.047	0.120
	(0.007)	(0.021)	(0.026)	(0.018)	(0.025)	(0.019)
$\text{experience}^2/100$	−0.224	−0.228	−0.080	−0.231	−0.032	−0.231
	(0.032)	(0.035)	(0.133)	(0.037)	(0.127)	(0.037)
Black	−0.190	−0.131	−0.103	−0.102	−0.064	−0.099
	(0.017)	(0.051)	(0.075)	(0.044)	(0.061)	(0.045)
south	−0.125	−0.105	−0.098	−0.095	−0.086	−0.094
	(0.015)	(0.023)	(0.0284)	(0.022)	(0.026)	(0.022)
urban	0.161	0.131	0.108	0.116	0.083	0.115
	(0.015)	(0.030)	(0.049)	(0.026)	(0.041)	(0.027)
Sargan				0.82	0.52	0.82
p-value				0.37	0.47	0.37

An IV Example: College Proximity

1. IV(a) uses college as an instrument for education.
2. IV(b) uses college, age, and age $^2/100$ as instruments for education, experience, and experience $^2/100$.
3. 2SLS(a) uses public and private as instruments for education.
4. 2SLS(b) uses public, private, age, and age ^{2} as instruments for education, experience, and experience $^2/100$.
5. LIML uses public and private as instruments for education.

Reduced Form

- The **reduced form** is the relationship between the endogenous variables and the instruments Z.
- A linear reduced form model for Y_2 is

$$Y_2 = \Gamma' Z + u_2 = \Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2$$

- The $\ell \times k_2$ coefficient matrix Γ is defined by **linear projection**:

$$\Gamma = \mathbb{E}\left[ZZ'\right]^{-1} \mathbb{E}\left[ZY_2'\right]$$

- **This implies $\mathbb{E}[Zu_2'] = 0$.**

Reduced Form

- We also construct the reduced form for Y_1.

$$\begin{aligned} Y_1 &= Z_1'\beta_1 + \left(\Gamma_{12}'Z_1 + \Gamma_{22}'Z_2 + u_2\right)'\beta_2 + e \\ &= Z_1'\lambda_1 + Z_2'\lambda_2 + u_1 \\ &= Z'\lambda + u_1 \end{aligned}$$

where

$$\begin{aligned} \lambda_1 &= \beta_1 + \Gamma_{12}\beta_2 \\ \lambda_2 &= \Gamma_{22}\beta_2 \\ u_1 &= u_2'\beta_2 + e \end{aligned}$$

Reduced Form

- We can also write
$$\lambda = \bar{\Gamma}\beta$$
where
$$\bar{\Gamma} = \begin{bmatrix} \boldsymbol{I}_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{k_1} & \Gamma \\ 0 & \end{bmatrix}$$
- Together, the reduced form equations for the system are
$$Y_1 = \lambda' Z + u_1$$
$$Y_2 = \Gamma' Z + u_2$$
- The reduced form equations are projections so the coefficients may be estimated by least squares.
$$\widehat{\Gamma} = \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \left(\sum_{i=1}^{n} Z_i Y_{2i}' \right)$$
$$\widehat{\lambda} = \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \left(\sum_{i=1}^{n} Z_i Y_{1i} \right)$$

Identification

- A parameter is identified if it is a unique function of the probability distribution of the observables.

$$\Gamma = \mathbb{E}\left[ZZ'\right]^{-1} \mathbb{E}\left[ZY_2'\right]$$
$$\lambda = \mathbb{E}\left[ZZ'\right]^{-1} \mathbb{E}\left[ZY_1\right]$$

- We are interested in the structural parameter β. It relates to (λ, Γ) through $\lambda = \bar{\Gamma}\beta$. β is identified if it uniquely determined by this relation.
- This is a set of ℓ equations with k unknowns with $\ell \geq k$. From linear algebra we know that there is a unique solution if and only if $\bar{\Gamma}$ has full rank k.

$$\text{rank}(\bar{\Gamma}) = k$$

Identification

- We can write $\bar{\Gamma} = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']$.
- Combining this with $\lambda = \bar{\Gamma}\beta$ we obtain

$$\mathbb{E}\left[ZZ'\right]^{-1}\mathbb{E}[ZY_1] = \mathbb{E}\left[ZZ'\right]^{-1}\mathbb{E}\left[ZX'\right]\beta$$

or

$$\mathbb{E}[ZY_1] = \mathbb{E}\left[ZX'\right]\beta$$

which is a set of ℓ equations with k unknowns. This has a unique solution if (and only if)

$$\text{rank}\left(\mathbb{E}\left[ZX'\right]\right) = k$$

Instrumental Variables Estimator

- We consider the special case where the model is just-identified so that $\ell = k$.
- The assumption that Z is an instrumental variable implies that $\mathbb{E}[Ze] = 0$.
- Making the substitution $e = Y_1 - X'\beta$ we find $\mathbb{E}\left[Z\left(Y_1 - X'\beta\right)\right] = 0$. Expanding,

$$\mathbb{E}\left[ZY_1\right] - \mathbb{E}\left[ZX'\right]\beta = 0$$

- This is a system of $\ell = k$ equations and k unknowns. So

$$\beta = \left(\mathbb{E}\left[ZX'\right]\right)^{-1}\mathbb{E}\left[ZY_1\right]$$

- The instrumental variables (IV) estimator β replaces population by sample moments. We find

$$\widehat{\beta}_{\mathrm{iv}} = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Y_{1i}\right) = \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} Z_i Y_{1i}\right)$$

Instrumental Variables Estimator

- Alternatively, recall that when $\ell = k$, $\beta = \bar{\Gamma}^{-1}\lambda$.
- Replacing $\bar{\Gamma}$ and λ by their least squares estimators, we can construct what is called the **Indirect Least Squares (ILS) estimator**. Using the matrix algebra representations

$$
\begin{aligned}
\widehat{\beta}_{\mathrm{ils}} = \widehat{\bar{\Gamma}}^{-1}\widehat{\lambda} &= \left(\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1}\left(\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Y}_1\right)\right) \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Y}_1\right) \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Y}_1\right)
\end{aligned}
$$

- We see that this equals the IV estimator. Thus the ILS and IV estimators are identical.
- Given the IV estimator we define the residual $\widehat{e}_i = Y_{1i} - X_i'\widehat{\beta}_{\mathrm{iv}}$. It satisfies

$$
\boldsymbol{Z}'\widehat{\boldsymbol{e}} = \boldsymbol{Z}'\boldsymbol{Y}_1 - \boldsymbol{Z}'\boldsymbol{X}\left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Y}_1\right) = 0
$$

Demeaned Representation

- Write the linear projection equation in the format $Y_1 = X'\beta + \alpha + e$ where α is the intercept and X does not contain a constant.
- Similarly, partition the instrument as $(1, Z)$ where Z does not contain a constant. We can write the IV estimator for the i^{th} equation as

$$Y_{1i} = X_i'\widehat{\beta}_{\mathrm{iv}} + \widehat{\alpha}_{\mathrm{iv}} + \widehat{e}_i$$

- The orthogonality condition implies the two-equation system

$$\sum_{i=1}^{n} \left(Y_{1i} - X_i'\widehat{\beta}_{\mathrm{iv}} - \widehat{\alpha}_{\mathrm{iv}} \right) = 0$$
$$\sum_{i=1}^{n} Z_i \left(Y_{1i} - X_i'\widehat{\beta}_{\mathrm{iv}} - \widehat{\alpha}_{\mathrm{iv}} \right) = 0$$

Demeaned Representation

- The first equation implies $\widehat{\alpha}_{\mathrm{iv}} = \bar{Y}_1 - \bar{X}'\widehat{\beta}_{\mathrm{iv}}$. Substituting into the second equation

$$\sum_{i=1}^{n} Z_i \left(\left(Y_{1i} - \bar{Y}_1 \right) - \left(X_i - \bar{X} \right)' \widehat{\beta}_{\mathrm{iv}} \right)$$

and solving for $\widehat{\beta}_{\mathrm{iv}}$ we find

$$\begin{aligned} \widehat{\beta}_{\mathrm{iv}} &= \left(\sum_{i=1}^{n} Z_i \left(X_i - \bar{X} \right)' \right)^{-1} \left(\sum_{i=1}^{n} Z_i \left(Y_{1i} - \bar{Y}_1 \right) \right) \\ &= \left(\sum_{i=1}^{n} \left(Z_i - \bar{Z} \right) \left(X_i - \bar{X} \right)' \right)^{-1} \left(\sum_{i=1}^{n} \left(Z_i - \bar{Z} \right) \left(Y_{1i} - \bar{Y}_1 \right) \right) \end{aligned}$$

Wald Estimator

- In many cases including the Card proximity example the excluded instrument is a binary (dummy) variable.
- Let's focus on that case and suppose that the model has just one endogenous regressor and no other regressors beyond the intercept.
- The model can be written as $Y = X\beta + \alpha + e$ with $\mathbb{E}[e \mid Z] = 0$ and Z binary.

Wald Estimator

- Take expectations of the structural equation given $Z = 1$ and $Z = 0$, respectively. We obtain
$$\mathbb{E}[Y \mid Z = 1] = \mathbb{E}[X \mid Z = 1]\beta + \alpha$$
$$\mathbb{E}[Y \mid Z = 0] = \mathbb{E}[X \mid Z = 0]\beta + \alpha$$
- Subtracting and dividing we obtain an expression for the slope coefficient
$$\beta = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[X \mid Z = 1] - \mathbb{E}[X \mid Z = 0]}$$

Wald Estimator

- The natural moment estimator replaces the expectations by the averages within the "grouped data" where $Z_i = 1$ and $Z_i = 0$, respectively. That is, define the group means

$$\bar{Y}_1 = \frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i}, \quad \bar{Y}_0 = \frac{\sum_{i=1}^n (1 - Z_i) Y_i}{\sum_{i=1}^n (1 - Z_i)}$$
$$\bar{X}_1 = \frac{\sum_{i=1}^n Z_i X_i}{\sum_{i=1}^n Z_i}, \quad \bar{X}_0 = \frac{\sum_{i=1}^n (1 - Z_i) X_i}{\sum_{i=1}^n (1 - Z_i)}$$

  and the moment estimator

$$\widehat{\beta} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

- This is the "Wald estimator" of Wald (1940).

Wald Estimator

- The expression may appear like a distinct estimator from the IV estimator $\widehat{\beta}_{\text{iv}}$ but it turns out that they are the same.
- To see this, we know

$$\widehat{\beta}_{\mathrm{iv}} = \frac{\sum_{i=1}^{n} Z_i \left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n} Z_i \left(X_i - \bar{X}\right)} = \frac{\bar{Y}_1 - \bar{Y}}{\bar{X}_1 - \bar{X}}$$

- Then notice

$$\bar{Y}_1 - \bar{Y} = \bar{Y}_1 - \left(\frac{1}{n} \sum_{i=1}^{n} Z_i \bar{Y}_1 + \frac{1}{n} \sum_{i=1}^{n} (1 - Z_i) \bar{Y}_0 \right) = (1 - \bar{Z}) \left(\bar{Y}_1 - \bar{Y}_0\right)$$

and similarly

$$\bar{X}_1 - \bar{X} = (1 - \bar{Z}) \left(\bar{X}_1 - \bar{X}_0\right)$$

and hence

$$\widehat{\beta}_{\mathrm{iv}} = \frac{(1 - \bar{Z}) \left(\bar{Y}_1 - \bar{Y}_0\right)}{(1 - \bar{Z}) \left(\bar{X}_1 - \bar{X}_0\right)} = \widehat{\beta}$$

Two-Stage Least Squares

- The IV estimator described in the previous slides presumed $\ell = k$. Now we allow the general case of $\ell \geq k$.
- Examining the reduced-form equation in page 12 we see

$$
\begin{aligned}
Y_1 &= Z'\bar{\Gamma}\beta + u_1 \\
\mathbb{E}\left[Zu_1\right] &= 0
\end{aligned}
$$

- Defining $W = \bar{\Gamma}'Z$ we can write this as

$$
\begin{aligned}
Y_1 &= W'\beta + u_1 \\
\mathbb{E}\left[Wu_1\right] &= 0.
\end{aligned}
$$

- Suppose that $\bar{\Gamma}$ were known. Then we would estimate β by least squares of Y_1 on $W = \bar{\Gamma}'Z$

$$
\widehat{\beta} = \left(\boldsymbol{W}'\boldsymbol{W}\right)^{-1}\left(\boldsymbol{W}'\boldsymbol{Y}\right) = \left(\bar{\Gamma}'\boldsymbol{Z}'\boldsymbol{Z}\bar{\Gamma}\right)^{-1}\left(\bar{\Gamma}'\boldsymbol{Z}'\boldsymbol{Y}_1\right)
$$

Two-Stage Least Squares

- While this is infeasible we can estimate $\bar{\Gamma}$ from the reduced form regression.
- Replacing $\bar{\Gamma}$ with its estimator $\widehat{\Gamma} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{X})$ we obtain

$$
\begin{aligned}
\widehat{\beta}_{\text{2sls}} &= \left(\widehat{\Gamma}'\boldsymbol{Z}'\boldsymbol{Z}\widehat{\Gamma}\right)^{-1}\left(\widehat{\Gamma}'\boldsymbol{Z}'\boldsymbol{Y}_1\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-\mathbf{1}}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y}_1 \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y}_1
\end{aligned}
$$

- This is called the two-stage-least squares (2SLS) estimator.
- You can show that if the model is just-identified, so that $k = \ell$, then 2SLS simplifies to the IV estimator.

Two-Stage Least Squares

- A useful representation of 2SLS arises if we define the projection matrix

$$\boldsymbol{P_Z} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'$$

- Because $\boldsymbol{P_Z}$ is idempotent we can also write the 2SLS estimator as

$$\widehat{\beta}_{\text{2sls}} = \left(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{P_Z}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{Y}_1 = \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{Y}_1$$

  which is the least squares estimator obtained by regressing Y_1 on the fitted values $\widehat{X}$.
- This is the source of the "two-stage" name as it can be computed as follows.
  1. Regress X on Z to obtain the fitted $\widehat{X}$: $\widehat{\Gamma} = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{X}\right)$ and $\widehat{\boldsymbol{X}} = \boldsymbol{Z}\widehat{\Gamma} = \boldsymbol{P_Z}\boldsymbol{X}$.
  2. Regress Y_1 on $\widehat{X}$: $\widehat{\beta}_{\text{2sls}} = \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{Y}_1$.

Consistency of 2SLS

Assumptions:

1. The variables $\left(Y_{1i}, X_i, Z_i\right)$, $i = 1, \ldots, n$, are independent and identically distributed.
2. $\mathbb{E}\left[Y_1^2\right] < \infty$
3. $\mathbb{E}\|X\|^2 < \infty$
4. $\mathbb{E}\|Z\|^2 < \infty$
5. $\mathbb{E}\left[ZZ'\right]$ is positive definite.
6. $\mathbb{E}\left[ZX'\right]$ has full rank k.
7. $\mathbb{E}[Ze] = 0$

Under those assumptions,

$$\widehat{\beta}_{\mathrm{2sls}} \xrightarrow[p]{} \beta \text{ as } n \to \infty.$$

Consistency of 2SLS

- The proof of this consistency result is similar to that for least squares.

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

- This separates out the stochastic component. Re-writing and applying the WLLN and CMT

$$\begin{aligned}\widehat{\beta}_{2\text{sls}} - \beta &= \left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1}\\ &\times \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{e}\right)\\ &\underset{p}{\longrightarrow} \left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\mathbb{E}[Ze] = 0\end{aligned}$$

Asymptotic Distribution of 2SLS

In additions to assumptions in page 27, we further assume:

1. $\mathbb{E}\left[Y_1^4\right] < \infty$
2. $\mathbb{E}\|X\|^4 < \infty$.
3. $\mathbb{E}\|Z\|^4 < \infty$.
4. $\Omega = \mathbb{E}\left[ZZ'e^2\right]$ is positive definite.

Under those assumptions, as $n \to \infty$.

$$\sqrt{n}\left(\widehat{\beta}_{\text{2sls}} - \beta\right) \xrightarrow[d]{} \mathrm{N}\left(0, \boldsymbol{V}_\beta\right)$$

where

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

Asymptotic Distribution of 2SLS

- We have the equation

$$\sqrt{n}\left(\widehat{\beta}_{\mathrm{2sls}}-\beta\right)=\left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1}$$
$$*\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{e}\right)$$

- We apply the WLLN and CMT for the moment matrices involving X and Z the same as in the proof of consistency. In addition, by the CLT for i.i.d. observations

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{e}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_ie_i\xrightarrow[d]{}\mathrm{N}(0,\Omega)$$

because the vector Z_ie_i is i.i.d. and mean zero, and has a finite second moment. We obtain

$$\sqrt{n}\left(\widehat{\beta}_{\mathrm{2sls}}-\beta\right)\xrightarrow[d]{}\left(\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{Q}_{ZZ}^{-1}\,\mathrm{N}(0,\Omega)=\mathrm{N}\left(0,\boldsymbol{V}_\beta\right)$$

Local Average Treatment Effects

- In a pair of influential papers, Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) proposed an new interpretation of the instrumental variables estimator using the potential outcomes model.
- We will restrict attention to the case that the endogenous regressor X and excluded instrument Z are binary variables.
- We write the model as a pair of potential outcome functions. The dependent variable Y is a function of the regressor and an unobservable vector U, $Y = h(X, U)$, and the endogenous regressor X is a function of the instrument Z and U, $X = g(Z, U)$.
- **U and Z are independent.**

Local Average Treatment Effects

- In this framework the outcomes are determined by the random vector U and the exogenous instrument Z. This determines X which determines Y.
- To put this in the context of the college proximity example the variable U is everything specific about an individual. Given college proximity Z the person decides to attend college or not. The person's wage is determined by the individual attributes U as well as college attendance X but is not directly affected by college proximity Z.
- We can omit the random variable U from the notation as follows. An individual has a realization U. We then set $Y(x) = h(x, U)$ and $X(z) = g(z, U)$. Also, given a realization Z the observables are $X = X(Z)$ and $Y = Y(X)$.

Local Average Treatment Effects

- In this model the causal effect of college for an individual is $C = Y(1) - Y(0)$. **This is individual-specific and random.**
- We would like to learn about the distribution of the causal effects, or at least features of the distribution.
- A common feature of interest is the average treatment effect (ATE)

$$\mathrm{ATE} = \mathbb{E}[C] = \mathbb{E}[Y(1) - Y(0)]$$

- This, however, it typically not feasible to estimate allowing for endogenous X without strong assumptions.

Local Average Treatment Effects

- One particular feature of interest emphasized by Imbens and Angrist (1994) is the local average treatment effect (LATE).
- Roughly, this is **the average effect upon those effected by the instrumental variable.**
- Consider the college proximity example. In the potential outcomes framework each person is fully characterized by their individual unobservable U. Given U, their decision to attend college is a function of the proximity indicator Z.
- For some students, proximity has no effect on their decision. For other students, it has an effect in the specific sense that given $Z = 1$ they choose to attend college while if $Z = 0$ they choose to not attend. We can summarize the possibilities with the following chart which is based on labels developed by Angrist, Imbens and Rubin (1996).

Local Average Treatment Effects

	$X(0) = 0$	$X(0) = 1$
$X(1) = 0$	Never Takers	Defiers
$X(1) = 1$	Compliers	Always Takers

- We need to assume that there are no Defiers, or equivalently that $X(1) \geq X(0)$ - a "**monotonicity**" condition.
- We can distinguish the types in the table by the relative values of $X(1) - X(0)$. For Never-Takers and Always-Takers $X(1) - X(0) = 0$, while for Compliers $X(1) - X(0) = 1$.

Local Average Treatment Effects

- We are interested in the treatment effect $C = h(1, U) - h(0, U)$ of college on wages. The average treatment effect (ATE) is its expectation $\mathbb{E}[Y(1) - Y(0)]$.
- To estimate the ATE we need observations of both $Y(0)$ and $Y(1)$ which means we need to observe some individuals who attend college and some who do not attend college.
- Consider the group "Never-Takers". They never attend college so we only observe $Y(0)$. It is thus impossible to estimate the ACE of college for this group.
- Similarly consider the group "Always-Takers". They always attend college so we only observe $Y(1)$ and again we cannot estimate the ACE of college for this group.
- The group for which we can estimate the ACE are the "Compliers". The ACE for this group is

$$\mathrm{LATE} = \mathbb{E}[Y(1) - Y(0) \mid X(1) > X(0)]$$

Local Average Treatment Effects

- Interestingly, we will show below that

$$\text{LATE} = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[X \mid Z = 1] - \mathbb{E}[X \mid Z = 0]}$$

- That is, LATE equals the Wald expression for the slope coefficient in the IV regression model.
- This means that **when treatment effects are potentially heterogeneous we can interpret IV as an estimator of LATE**.
- The equality occurs under the following conditions: **U and Z are independent and $\mathbb{P}[X(1) - X(0) < 0] = 0$.**

Local Average Treatment Effects

- One interesting feature about LATE is that its value can depend on the instrument Z and the distribution of causal effects C in the population.
- Suppose that instead of the Card proximity instrument we consider an instrument based on the financial cost of local college attendance.
- Some students may be responsive to proximity but not finances, and conversely.
- If the causal effect C has a different average in these two groups of students then LATE will be different when calculated with these two instruments.
- Why? Think about the difference between standard IV setting and the potential outcome setting.

Local Average Treatment Effects

- The realized value of X can be written as

$$X = (1 - Z)X(0) + ZX(1) = X(0) + Z(X(1) - X(0))$$

  Similarly, $Y = Y(0) + X(Y(1) - Y(0)) = Y(0) + XC$
- Combining,

$$Y = Y(0) + X(0)C + Z(X(1) - X(0))C$$

- The independence of u and Z implies independence of $(Y(0), Y(1), X(0), X(1), C)$ and Z. Thus

$$\mathbb{E}[Y \mid Z = 1] = \mathbb{E}[Y(0)] + \mathbb{E}[X(0)C] + \mathbb{E}[(X(1) - X(0))C]$$

  and

$$\mathbb{E}[Y \mid Z = 0] = \mathbb{E}[Y(0)] + \mathbb{E}[X(0)C]$$

Local Average Treatment Effects

- Subtracting we obtain

$$
\begin{aligned}
&\mathbb{E}[Y \mid Z=1]-\mathbb{E}[Y \mid Z=0]=\mathbb{E}[(X(1)-X(0))C] \\
&=1 \times \mathbb{E}[C \mid X(1)-X(0)=1]\mathbb{P}[X(1)-X(0)=1] \\
&+0 \times \mathbb{E}[C \mid X(1)-X(0)=0]\mathbb{P}[X(1)-X(0)=0] \\
&+(-1) \times \mathbb{E}[C \mid X(1)-X(0)=-1]\mathbb{P}[X(1)-X(0)=-1] \\
&=\mathbb{E}[C \mid X(1)-X(0)=1](\mathbb{E}[X \mid Z=1]-\mathbb{E}[X \mid Z=0])
\end{aligned}
$$

where the final equality uses $\mathbb{P}[X(1)-X(0)<0]=0$ and

$$\mathbb{P}[X(1)-X(0)=1]=\mathbb{E}[X(1)-X(0)]=\mathbb{E}[X \mid Z=1]-\mathbb{E}[X \mid Z=0]$$

- Rearranging

$$\text{LATE} =\mathbb{E}[C \mid X(1)-X(0)=1]=\frac{\mathbb{E}[Y \mid Z=1]-\mathbb{E}[Y \mid Z=0]}{\mathbb{E}[X \mid Z=1]-\mathbb{E}[X \mid Z=0]}$$

Endogeneity Tests

- The structural and reduced form equations for the standard IV model are

$$\begin{aligned} Y &= X_1'\beta_1 + X_2'\beta_2 + e \\ X_2 &= \Gamma_{12}'Z_1 + \Gamma_{22}'Z_2 + u_2 \end{aligned}$$

- The 2SLS estimator allows the regressor X_2 to be endogenous, meaning that X_2 is correlated with the structural error e.
- If this correlation is zero then X_2 is exogenous and the structural equation can be estimated by least squares.
- This is a testable restriction. Effectively, the null hypothesis is

$$\mathbb{H}_0 : \mathbb{E}\left[X_2 e\right] = 0$$

with the alternative

$$\mathbb{H}_1 : \mathbb{E}\left[X_2 e\right] \neq 0$$

Endogeneity Tests

- Since the instrumental variable assumption specifies that $\mathbb{E}[Ze] = 0, X_2$ is endogenous (correlated with e) if u_2 and e are correlated. We can therefore consider the linear projection of e on u_2

$$
\begin{aligned}
e &= u_2'\alpha + v \\
\alpha &= \left(\mathbb{E}\left[u_2 u_2'\right]\right)^{-1} \mathbb{E}\left[u_2 e\right] \\
\mathbb{E}\left[u_2 v\right] &= 0.
\end{aligned}
$$

- Substituting this into the structural form equation we find

$$
\begin{aligned}
Y &= X_1'\beta_1 + X_2'\beta_2 + u_2'\alpha + v \\
\mathbb{E}\left[X_1 v\right] &= 0 \\
\mathbb{E}\left[X_2 v\right] &= 0 \\
\mathbb{E}\left[u_2 v\right] &= 0
\end{aligned}
$$

Endogeneity Tests

- Notice that $\mathbb{E}\left[X_2 e\right] = 0$ if and only if $\mathbb{E}\left[u_2 e\right] = 0$, so the hypothesis can be restated as $\mathbb{H}_0 : \alpha = 0$ against $\mathbb{H}_1 : \alpha \neq 0$.
- The null hypothesis can be tested using Durbin-Wu-Hausman tests, Wu-Hausman tests, or Hausman tests, depending on the author.

OverIdentification Tests

- When $\ell > k$ the model is overidentified meaning that there are more moments than free parameters. This is a restriction and is testable. Such tests are called overidentification tests.
- The instrumental variables model specifies $\mathbb{E}[Ze] = 0$. Equivalently, since $e = Y - X'\beta$ this is
$$\mathbb{E}[ZY] - \mathbb{E}\left[ZX'\right]\beta = 0$$
- This is an $\ell \times 1$ vector of restrictions on the moment matrices $\mathbb{E}[ZY]$ and $\mathbb{E}[ZX']$. Yet since β is of dimension k which is less than ℓ it is not certain if indeed such a β exists.

OverIdentification Tests

- Suppose there is a single endogenous regressor X_2, no X_1, and two instruments Z_1 and Z_2. Then the model specifies that

$$\mathbb{E}([Z_1 Y] = \mathbb{E}[Z_1 X_2]\beta$$

and

$$\mathbb{E}[Z_2 Y] = \mathbb{E}[Z_2 X_2]\beta$$

- It is like we are estimating by IV using just the instrument Z_1 or instead just using the instrument Z_2.
- If the overidentification hypothesis is correct: both are estimating the same parameter and both are consistent for β.
- If the overidentification hypothesis is false: the two estimators will converge to different probability limits and it is unclear if either probability limit is interesting.
- The overidentification hypothesis can be tested by the Sargan statistic test.

Weak Instruments

- When the reduced form coefficients are treated as weak, which means Γ is small, the 2SLS estimators are inconsistent.

$$\widehat{\beta}_{\mathrm{2sls}} - \beta = \left(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{e}\right)$$
$$\xrightarrow[d]{} \left(\left(\boldsymbol{Q}_Z\boldsymbol{C} + \xi_2\right)'\boldsymbol{Q}_Z^{-1}\left(\boldsymbol{Q}_Z\boldsymbol{C} + \xi_2\right)\right)^{-1}\left(\boldsymbol{Q}_Z\boldsymbol{C} + \xi_2\right)'\boldsymbol{Q}_Z^{-1}\xi_e$$

- It is asymptotically random, and its asymptotic distribution is non-normal.
- Weak IV is tested by Stock and Yogo (2005).
- Rule of Numb: the first-stage F statistic should exceed 10 in order to achieve reliable IV inference for the case of one instrument.