International School of Economics at TSU
Econometrics II
Lasha Chochua

Problem Set 2

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1

Define the **2SLS residuals**:

$$\hat{\mathbf{e}} = \mathbf{Y}_1 - \mathbf{X}\hat{\beta}_{\text{2sls}}$$

Define the **fitted values** from the first stage:

$$\hat{\mathbf{X}} = \mathbf{P}_Z\mathbf{X} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$$

Show that:

$$\hat{\mathbf{X}}'\hat{\mathbf{e}} = 0$$

Hint: Use the definition of $\hat{\beta}_{\text{2sls}}$ and the fact that $\hat{\mathbf{X}}' = \mathbf{X}'\mathbf{P}_Z$ (since $\mathbf{P}_Z$ is symmetric).

Solution

Orthogonality of 2SLS Fitted Values

- Substitute the definition of $\hat{\mathbf{e}}$:

$$\hat{\mathbf{X}}'\hat{\mathbf{e}} = \hat{\mathbf{X}}'(\mathbf{Y}_1 - \mathbf{X}\hat{\beta}_{\text{2sls}}) = \hat{\mathbf{X}}'\mathbf{Y}_1 - \hat{\mathbf{X}}'\mathbf{X}\hat{\beta}_{\text{2sls}}$$

- Write $\hat{\mathbf{X}}' = \mathbf{X}'\mathbf{P}_Z$:

$$= \mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1 - \mathbf{X}'\mathbf{P}_Z\mathbf{X}\hat{\beta}_{\text{2sls}}$$

- Recall the 2SLS estimator:

$$\hat{\beta}_{\text{2sls}} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1$$

- Therefore $\mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1 = \mathbf{X}'\mathbf{P}_Z\mathbf{X}\hat{\beta}_{\text{2sls}}$, and:

$$\hat{\mathbf{X}}'\hat{\mathbf{e}} = \mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1 - \mathbf{X}'\mathbf{P}_Z\mathbf{X}\hat{\beta}_{\text{2sls}} = 0 \qquad \blacksquare$$

- **Contrast with OLS**: OLS residuals satisfy $\mathbf{X}'\hat{\mathbf{e}} = 0$ – here it is $\hat{\mathbf{X}}'$ (projected regressors), not $\mathbf{X}'$, that is orthogonal.
- **Implication**: when overidentified ($\ell > k$), $\mathbf{Z}'\hat{\mathbf{e}} \neq 0$ in general – only the k-dimensional projected subspace is guaranteed orthogonal to $\hat{\mathbf{e}}$.

Problem 2

$$y_i = \mathbf{x}_i'\beta + e_i \quad (1)$$
$$\mathbb{E}(\mathbf{z}_i e_i) = 0 \quad (2)$$

The dimensions are: $\mathbf{x}_i$, $\mathbf{z}_i$, and β are $k \times 1$, $k > 1$, and y_i and e_i are 1×1. Let

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xz} \\ \mathbf{Q}_{zx} & \mathbf{Q}_{zz} \end{bmatrix} = \begin{bmatrix} \mathbb{E}(\mathbf{x}_i\mathbf{x}_i') & \mathbb{E}(\mathbf{x}_i\mathbf{z}_i') \\ \mathbb{E}(\mathbf{z}_i\mathbf{x}_i') & \mathbb{E}(\mathbf{z}_i\mathbf{z}_i') \end{bmatrix}$$

Assume both $\mathbf{Q}_{xx}$ and $\mathbf{Q}_{zz}$ have full rank k.

Let $\hat{\beta}$ be the least-squares estimate obtained by regressing y_i on $\mathbf{x}_i$, and let $\tilde{\beta}$ be the 2SLS estimator obtained by estimation of (1) using the instrument $\mathbf{z}_i$.

1. Find

$$\delta = \operatorname*{plim}_{n\to\infty} \left(\hat{\beta} - \tilde{\beta}\right)$$

Solution:

$$\hat{\beta} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i e_i\right) \xrightarrow{p} \beta + \mathbf{Q}_{xx}^{-1}\mathbb{E}(\mathbf{x}_i e_i)$$

Because the equation is just-identified,

$$
\begin{aligned}
\tilde{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i y_i\right) \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i e_i\right) \\
&\xrightarrow{p} \beta + \mathbf{Q}_{zx}^{-1}\cdot 0 = \beta
\end{aligned}
$$

Thus

$$
\begin{aligned}
\delta &= \operatorname*{plim}_{n\to\infty}\left(\hat{\beta}-\tilde{\beta}\right) \\
&= \beta + \mathbf{Q}_{xx}^{-1}\mathbb{E}(\mathbf{x}_i e_i) - \beta \\
&= \mathbf{Q}_{xx}^{-1}\mathbb{E}(\mathbf{x}_i e_i)
\end{aligned}
$$

2. Suppose that in addition to (1) and (2),

$$
\mathbb{E}(\mathbf{x}_i e_i) = 0 \quad (3)
$$

Quite simply, what does this condition mean? What is δ under this assumption?

Solution:

Equation (3) means that $\mathbf{x}_i$ is exogenous. Under this assumption, $\delta = 0$.

3. Write the difference $\hat{\beta}-\tilde{\beta}$ as a function of sample moments of $\mathbf{x}_i$, $\mathbf{z}_i$, and e_i.

Solution:

Differencing the above equations,

$$
\begin{aligned}
\hat{\beta}-\tilde{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i e_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i e_i\right) \\
&= \left(\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} - \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{x}_i\\ \mathbf{z}_i\end{pmatrix} e_i\right)
\end{aligned}
$$

4. Under (1)-(3), find the asymptotic distribution of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{pmatrix} e_i$$

as $n \to \infty$.

Solution:

Since $\mathbb{E}(\mathbf{x}_i e_i) = 0$, $\mathbb{E}(\mathbf{z}_i e_i) = 0$ and

$$\mathbb{E}\left(\begin{pmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{pmatrix} \begin{pmatrix} \mathbf{x}_i' & \mathbf{z}_i' \end{pmatrix} e_i^2 \right) = \begin{bmatrix} \mathbb{E}(\mathbf{x}_i \mathbf{x}_i' e_i^2) & \mathbb{E}(\mathbf{x}_i \mathbf{z}_i' e_i^2) \\ \mathbb{E}(\mathbf{z}_i \mathbf{x}_i' e_i^2) & \mathbb{E}(\mathbf{z}_i \mathbf{z}_i' e_i^2) \end{bmatrix} = \begin{bmatrix} \Omega_{xx} & \Omega_{xz} \\ \Omega_{zx} & \Omega_{zz} \end{bmatrix} = \Omega,$$

say, then by the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{pmatrix} e_i \xrightarrow{d} \mathcal{N}(0, \Omega)$$

5. Under (1)-(3), find the asymptotic distribution of

$$\sqrt{n}\left(\hat{\beta} - \tilde{\beta}\right) \quad \text{as } n \to \infty.$$

Solution:

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \tilde{\beta}) &= \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{x}_i' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{pmatrix} e_i \right) \\ &\xrightarrow{d} \begin{pmatrix} \mathbf{Q}_{xx}^{-1} & -\mathbf{Q}_{zx}^{-1} \end{pmatrix} \mathcal{N}(0, \Omega) \sim \mathcal{N}(0, V) \end{aligned}$$

where

$$\begin{aligned} V &= \begin{pmatrix} \mathbf{Q}_{xx}^{-1} & -\mathbf{Q}_{zx}^{-1} \end{pmatrix} \begin{bmatrix} \Omega_{xx} & \Omega_{xz} \\ \Omega_{zx} & \Omega_{zz} \end{bmatrix} \begin{pmatrix} \mathbf{Q}_{xx}^{-1} \\ -\mathbf{Q}_{xz}^{-1} \end{pmatrix} \\ &= \mathbf{Q}_{xx}^{-1} \Omega_{xx} \mathbf{Q}_{xx}^{-1} - \mathbf{Q}_{xx}^{-1} \Omega_{xz} \mathbf{Q}_{xz}^{-1} - \mathbf{Q}_{zx}^{-1} \Omega_{zx} \mathbf{Q}_{xx}^{-1} + \mathbf{Q}_{zx}^{-1} \Omega_{zz} \mathbf{Q}_{xz}^{-1} \end{aligned}$$

6. Suppose that

$$\mathbb{E}(e_i^2 \mid \mathbf{x}_i, \mathbf{z}_i) = \sigma^2 \tag{4}$$

How does the asymptotic variance from question 5 simplify under (4)?

Solution:

Under (4),

$$\begin{bmatrix} \Omega_{xx} & \Omega_{xz} \\ \Omega_{zx} & \Omega_{zz} \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xz} \\ \mathbf{Q}_{zx} & \mathbf{Q}_{zz} \end{bmatrix}$$

so

$$\begin{aligned} V &= \mathbf{Q}_{xx}^{-1}\Omega_{xx}\mathbf{Q}_{xx}^{-1} - \mathbf{Q}_{zx}^{-1}\Omega_{zx}\mathbf{Q}_{xx}^{-1} - \mathbf{Q}_{xx}^{-1}\Omega_{xz}\mathbf{Q}_{xz}^{-1} + \mathbf{Q}_{zx}^{-1}\Omega_{zz}\mathbf{Q}_{xz}^{-1} \\ &= \sigma^2 \left(\mathbf{Q}_{xx}^{-1} - \mathbf{Q}_{xx}^{-1} - \mathbf{Q}_{xx}^{-1} + \mathbf{Q}_{zx}^{-1}\mathbf{Q}_{zz}\mathbf{Q}_{xz}^{-1}\right) \\ &= \sigma^2 \left(\mathbf{Q}_{zx}^{-1}\mathbf{Q}_{zz}\mathbf{Q}_{xz}^{-1} - \mathbf{Q}_{xx}^{-1}\right) \end{aligned}$$

7. Propose an estimator of the asymptotic variance under (4).

Solution:

$$\hat{V} = \hat{\sigma}^2 \left(\hat{\mathbf{Q}}_{zx}^{-1}\hat{\mathbf{Q}}_{zz}\hat{\mathbf{Q}}_{xz}^{-1} - \hat{\mathbf{Q}}_{xx}^{-1}\right)$$

where

$$\hat{\mathbf{Q}}_{xx} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i', \quad \hat{\mathbf{Q}}_{xz} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{z}_i', \quad \hat{\mathbf{Q}}_{zz} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i', \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2, \quad \hat{e}_i = y_i - \mathbf{x}_i'\hat{\beta}.$$

8. Propose a test statistic for (3) under (4) and find its asymptotic distribution under the assumption that $\mathbf{Q} > 0$.

Solution:

A test for H_0 is

$$W_n = n\left(\hat{\beta} - \tilde{\beta}\right)' \hat{V}^{-1} \left(\hat{\beta} - \tilde{\beta}\right)$$

Let the distribution in question 5 be $N \sim \mathcal{N}(0, V)$. Then the asymptotic distribution of W_n is

$$W_n \xrightarrow{d} N'V^{-1}N \sim \chi^2_k$$

9. Describe how to use this statistic to test the hypothesis that $\mathbf{x}_i$ is exogenous.

Solution:

Under exogeneity, W_n is asymptotically χ^2_k. To test exogeneity, we compare W_n with the χ^2_k distribution.

If W_n is smaller than the 5% critical value, we do not reject the hypothesis of exogeneity.

If W_n is larger than the critical value, we reject exogeneity in favor of endogeneity.

The test works because under the alternative,

$$\widehat{\beta} - \widetilde{\beta} \xrightarrow{p} \beta^* - \beta = \mathbf{Q}_{xx}^{-1}\delta \neq 0,$$

so $W_n \xrightarrow{p} \infty$.

10. Show where $\mathbf{Q} > 0$ is used in the answer to question 8.

Solution:

The asymptotic distribution implicitly assumed

$$V = \sigma^2 \left(\mathbf{Q}_{zx}^{-1}\mathbf{Q}_{zz}\mathbf{Q}_{xz}^{-1} - \mathbf{Q}_{xx}^{-1} \right) > 0.$$

This is true iff

$$\mathbf{Q}_{xx}^{-1} < \mathbf{Q}_{xz}^{-1}\mathbf{Q}_{zz}\mathbf{Q}_{xz}^{-1}$$

or iff

$$\mathbf{Q}_{xx} > \left(\mathbf{Q}_{zx}^{-1}\mathbf{Q}_{zz}\mathbf{Q}_{xz}^{-1} \right)^{-1} = \mathbf{Q}_{xx}\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{zx}$$

or iff

$$\mathbf{Q}_{xx} - \mathbf{Q}_{xz}\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{zx} > 0$$

This holds when $V > 0$, but not generally.

Problem 3

Let the structural equation be:

$$Y_1 = X'\beta + e, \quad \mathbb{E}[Ze] = 0$$

with 2SLS estimator:

$$\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_1$$

Let $\mathbf{C}$ be an invertible $\ell \times \ell$ matrix and define $\tilde{Z} = \mathbf{C}Z$.

(a) Show that replacing Z with $\tilde{Z} = \mathbf{C}Z$ leaves $\hat{\beta}_{2SLS}$ unchanged. That is, show:

$$(\tilde{\mathbf{X}}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{Y}_1 = \hat{\beta}_{2SLS}$$

where $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{C}'$ is the $n \times \ell$ sample matrix of transformed instruments.

(b) Using the projection matrix $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, show that $\mathbf{P}_{\tilde{Z}} = \mathbf{P}_Z$. What does this say geometrically?

(c) Using (b), write the 2SLS estimator in the form $\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1$ and argue directly that invariance to $\mathbf{C}$ follows from $\mathbf{P}_{\tilde{Z}} = \mathbf{P}_Z$.

(d) Now show that the asymptotic variance $V_\beta^0 = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}\sigma^2$ is also invariant to replacing Z with $\tilde{Z} = \mathbf{C}Z$. Derive each step explicitly showing how the $\mathbf{C}$ matrices cancel.

(e) Does invariance hold if $\mathbf{C}$ is not invertible? Explain what goes wrong, relating your answer to the rank condition $\text{rank}(\mathbb{E}[ZX']) = k$.

Solution

(a) Note that $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{C}'$. Substitute directly:

$$\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} = \mathbf{C}\mathbf{Z}'\mathbf{Z}\mathbf{C}'$$

$$\tilde{\mathbf{Z}}'\mathbf{X} = \mathbf{C}\mathbf{Z}'\mathbf{X}$$

$$(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1} = (\mathbf{C}')^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}^{-1}$$

Now compute $\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{X}$:

$$= \mathbf{Z}\mathbf{C}'(\mathbf{C}')^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{Z}'\mathbf{X} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$$

Therefore:

$$\mathbf{X}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{X} = \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$$

and similarly $\mathbf{X}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{Y}_1 = \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_1$.

Hence $\hat{\beta}_{2SLS}$ is unchanged. ■

(b) Compute $\mathbf{P}_{\tilde{Z}}$ directly:

$$\mathbf{P}_{\tilde{Z}} = \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}' = \mathbf{Z}\mathbf{C}'(\mathbf{C}')^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{Z}' = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{P}_Z$$

Geometrically: $\mathbf{P}_Z$ projects onto the **column space** of $\mathbf{Z}$. Since $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{C}'$ and $\mathbf{C}$ is invertible, $\tilde{\mathbf{Z}}$ spans the **same column space** as $\mathbf{Z}$ – so the projection is identical. ■

(c) Writing 2SLS via the projection matrix:

$$\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1$$

Since $\mathbf{P}_{\tilde{Z}} = \mathbf{P}_Z$ from part (b), replacing Z with $\tilde{Z}$ gives:

$$(\mathbf{X}'\mathbf{P}_{\tilde{Z}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\tilde{Z}}\mathbf{Y}_1 = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y}_1 = \hat{\beta}_{2SLS}$$

Invariance follows directly because 2SLS depends on Z **only through $\mathbf{P}_Z$**, which is determined by the column space of Z, not its particular basis. ■

(d) Replacing Z with $\tilde{Z} = \mathbf{C}Z$ in population moments:

$$\tilde{Q}_{XZ} = \mathbb{E}[X\tilde{Z}'] = \mathbb{E}[XZ'\mathbf{C}'] = Q_{XZ}\mathbf{C}'$$
$$\tilde{Q}_{ZZ} = \mathbb{E}[\tilde{Z}\tilde{Z}'] = \mathbf{C}\mathbb{E}[ZZ']\mathbf{C}' = \mathbf{C}Q_{ZZ}\mathbf{C}'$$
$$\tilde{Q}_{ZX} = \mathbf{C}Q_{ZX}$$

Substitute into V_β^0:

$$\tilde{V}_\beta^0 = (\tilde{Q}_{XZ}\tilde{Q}_{ZZ}^{-1}\tilde{Q}_{ZX})^{-1}\sigma^2$$
$$= \left(Q_{XZ}\mathbf{C}'(\mathbf{C}')^{-1}Q_{ZZ}^{-1}\mathbf{C}^{-1}\mathbf{C}Q_{ZX}\right)^{-1}\sigma^2$$
$$= \left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\sigma^2 = V_\beta^0$$

The $\mathbf{C}$ matrices cancel exactly. ■

(e) If $\mathbf{C}$ is not invertible, then $\text{rank}(\tilde{\mathbf{Z}}) = \text{rank}(\mathbf{Z}\mathbf{C}') < \ell$. This means:

- The column space of $\tilde{\mathbf{Z}}$ is a **strict subset** of the column space of $\mathbf{Z}$
- $\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} = \mathbf{C}\mathbf{Z}'\mathbf{Z}\mathbf{C}'$ is **singular** – not invertible
- The relevance condition $\text{rank}(\mathbb{E}[\tilde{Z}X']) = \text{rank}(\mathbf{C}Q_{ZX}) = k$ fails since $\mathbf{C}$ is rank-deficient

So 2SLS is not identified with $\tilde{Z}$ – we have lost identifying variation by collapsing the instrument space. Invertibility of $\mathbf{C}$ is precisely what preserves the span of Z and hence identification.

Problem 4

Let $\theta = r(\beta) \in \mathbb{R}^q$ be a differentiable function of the 2SLS estimator, with $\hat{\theta} = r(\hat{\beta}_{2SLS})$. Denote $\mathbf{R} = \frac{\partial}{\partial \beta} r(\beta)'$ $(k \times q)$. We test $H_0 : \theta = \theta_0$ using the Wald statistic:

$$W = n(\hat{\theta} - \theta_0)' \hat{V}_\theta^{-1} (\hat{\theta} - \theta_0)$$

(a) Starting from Theorem 5 in the related slides:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\theta), \quad V_\theta = \mathbf{R}' V_\beta \mathbf{R}$$

Show step by step that $W \xrightarrow{d} \chi^2_q$ under H_0. You must explicitly use the definition of the χ^2 distribution.

(b) At which step does $\hat{V}_\theta \xrightarrow{p} V_\theta$ enter the argument? Which theorem justifies replacing V_θ with $\hat{V}_\theta$ without changing the limiting distribution?

(c) Now suppose $e_i \mid X_i, Z_i \sim \mathcal{N}(0, \sigma^2)$ (normality). Consider the special case $q = 1$ (scalar θ). Show that the t-statistic:

$$t = \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{\hat{V}_\theta}}$$

satisfies $t^2 = W$ and derive its exact distribution under normality. At which step does normality play a role?

(d) Now consider general $q \geq 1$ under normality. The F-statistic is defined as:

$$F = \frac{W}{q}$$

Show that $F \sim F(q, n - k_1 - 2k_2)$ by expressing W as a ratio of two independent χ^2 random variables, each divided by their degrees of freedom. State clearly what each χ^2 corresponds to.

(e) Why does the F distribution reduce to the asymptotic χ^2_q / q as $n \to \infty$? Show this using the definition $F(q, m) \xrightarrow{d} \chi^2_q / q$ as $m \to \infty$.

Solution

(a) We proceed in four steps.

1. Under H_0, by Theorem 5:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

2. Since V_θ is positive definite, factor $V_\theta = V_\theta^{1/2} V_\theta^{1/2}$ and multiply by $V_\theta^{-1/2}$:

$$V_\theta^{-1/2}\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_q)$$

This is a vector of q independent standard normals.

3. By definition of the χ^2 distribution: if $\xi \sim \mathcal{N}(0, I_q)$ then $\|\xi\|^2 = \xi'\xi \sim \chi^2_q$. Therefore:

$$\left\| V_\theta^{-1/2}\sqrt{n}(\hat{\theta} - \theta_0) \right\|^2 = n(\hat{\theta} - \theta_0)' V_\theta^{-1} (\hat{\theta} - \theta_0) \xrightarrow{d} \chi^2_q$$

4. Replace V_θ with $\hat{V}_\theta$ (justified in part b):

$$W = n(\hat{\theta} - \theta_0)' \hat{V}_\theta^{-1} (\hat{\theta} - \theta_0) \xrightarrow{d} \chi^2_q \qquad \blacksquare$$

(b) The replacement occurs at **Step 4**. We have:

$$W = n(\hat{\theta} - \theta_0)' \hat{V}_\theta^{-1} (\hat{\theta} - \theta_0)$$

Write this as $g(\sqrt{n}(\hat{\theta} - \theta_0), \hat{V}_\theta)$ where $g(a, B) = a'B^{-1}a$ is a continuous function. Since:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \xi \sim \mathcal{N}(0, V_\theta), \quad \hat{V}_\theta \xrightarrow{p} V_\theta$$

Slutsky's theorem justifies:

$$g(\sqrt{n}(\hat{\theta} - \theta_0), \hat{V}_\theta) \xrightarrow{d} g(\xi, V_\theta) = \xi' V_\theta^{-1} \xi \sim \chi^2_q$$

(c) When $q = 1$, V_θ is a scalar and $\hat{V}_\theta > 0$. The t-statistic satisfies:

$$t^2 = \frac{n(\hat{\theta} - \theta_0)^2}{\hat{V}_\theta} = W$$

Under normality $e_i \mid X_i, Z_i \sim \mathcal{N}(0, \sigma^2)$:

- $\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, V_\theta)$ **exactly** in finite samples (not just asymptotically)
- The residual sum of squares $\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2} \sim \chi^2_{n-k_1-2k_2}$ **exactly**
- These are independent by normality

Normality enters at the first bullet – it turns the asymptotic normal approximation into an **exact** finite-sample result. Therefore:

$$t = \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{\hat{V}_\theta}} \sim t(n - k_1 - 2k_2)$$

and $t^2 = W \sim F(1, n-k_1-2k_2)$, consistent with part (d). ■

(d) Under normality, decompose W as follows.

Numerator. Under H_0 and normality:

$$W = n(\hat{\theta}-\theta_0)'\hat{V}_\theta^{-1}(\hat{\theta}-\theta_0) \sim \chi^2_q$$

exactly in finite samples – this is a quadratic form in a normal vector. Dividing by q:

$$\frac{W}{q} \sim \frac{\chi^2_q}{q}$$

Denominator. The variance estimator relies on:

$$\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2} \sim \chi^2_{n-k_1-2k_2}$$

from the control function regression with $n-k_1-2k_2$ residual degrees of freedom. This enters $\hat{V}_\theta$ through $\hat{\sigma}^2 = \frac{1}{n}\hat{\mathbf{e}}'\hat{\mathbf{e}}$.

Independence. Under normality, the numerator (a function of $\hat{\theta}$) and denominator (a function of $\hat{\mathbf{e}}$) are independent by the normal theory analogue of the Gauss-Markov theorem.

Therefore by definition of the F distribution:

$$F = \frac{W/q}{\hat{\mathbf{e}}'\hat{\mathbf{e}}/(\sigma^2(n-k_1-2k_2))} = \frac{\chi^2_q/q}{\chi^2_{n-k_1-2k_2}/(n-k_1-2k_2)} \sim F(q,\, n-k_1-2k_2) \qquad \blacksquare$$

(e) By definition, if $A \sim \chi^2_q$ and $B_m \sim \chi^2_m$ independently, then:

$$F(q,m) = \frac{A/q}{B_m/m}$$

As $m \to \infty$, by LLN:

$$\frac{B_m}{m} \xrightarrow{p} 1$$

since $\mathbb{E}[B_m/m] = 1$ and $\mathrm{Var}(B_m/m) = 2/m \to 0$. Therefore by Slutsky:

$$F(q,m) = \frac{A/q}{B_m/m} \xrightarrow{d} \frac{A}{q} \sim \frac{\chi^2_q}{q}$$

In our case $m = n-k_1-2k_2 \to \infty$ as $n \to \infty$, so:

$$F = \frac{W}{q} \xrightarrow{d} \frac{\chi^2_q}{q}$$

which is equivalent to $W \xrightarrow{d} \chi^2_q$ – the asymptotic result. Without normality we can only use this limit; under normality the F distribution is **exact** for any n. ■

Problem 5

Part A – Intuition

- Recall the IV moment condition: $\mathbb{E}[\mathbf{Z}e] = 0$
- We have ℓ instruments but only k parameters to estimate
- Explain intuitively:
  1. Why overidentification ($\ell > k$) creates a **testable restriction**
  2. Why $\hat{\alpha} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}$ is the natural object to test
  3. Why the degrees of freedom of the test should be $\ell - k$, not ℓ

Part B – Derivation

- Under $H_0 : \mathbb{E}[\mathbf{Z}e] = 0$ and homoskedasticity $\mathbb{E}[e^2 \mid \mathbf{Z}] = \sigma^2$, show that:

$$S = \frac{\hat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{\ell-k}$$

- **Hints**:
  1. Write $\hat{\mathbf{e}} = \mathbf{e} - \mathbf{X}(\hat{\beta}_{\text{2sls}} - \beta)$ and show $\frac{1}{\sqrt{n}}\mathbf{Z}'\hat{\mathbf{e}} \xrightarrow{d} \frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}$ asymptotically
  2. Apply the CLT to $\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}$
  3. Count the moment conditions used up by estimation

Solution

Part A – Intuition

- We have ℓ moment conditions $\mathbb{E}[\mathbf{Z}e] = 0$ but only k free parameters in β
- Estimation uses exactly k of these conditions to pin down $\hat{\beta}_{\text{2sls}}$
- The remaining $\ell - k$ conditions are not used in estimation – they are left over and can be checked
- $\hat{\alpha} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}$ measures how much the residuals correlate with the instruments – under H_0 this should be close to zero
- There are ℓ instruments but k directions are already absorbed by estimating β, leaving $\ell - k$ free directions – hence degrees of freedom $= \ell - k$

Part B – Derivation

- Start from the definition $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\beta}_{\text{2sls}}$, add and subtract $\mathbf{X}\beta$, and use $\mathbf{Y} - \mathbf{X}\beta = \mathbf{e}$ from the structural model:

$$\hat{\mathbf{e}} = \mathbf{e} - \mathbf{X}(\hat{\beta}_{\text{2sls}} - \beta)$$

- As $n \to \infty$, $\hat{\beta}_{\text{2sls}} \to \beta$, so the correction term vanishes and $\hat{\mathbf{e}} \to \mathbf{e}$
- Therefore:

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

- Therefore:

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

- By the CLT for i.i.d. vectors:

$$\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e} \xrightarrow{d} \mathcal{N}(0,\, \sigma^2\mathbf{Q}_{ZZ})$$

- By consistency of 2SLS and LLN:

$$\sqrt{n}(\hat{\beta}_{\text{2sls}} - \beta) = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1} \cdot \frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e} + o_p(1)$$

- Define $\mathbf{A} = \mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}$. Then:

$$\frac{1}{\sqrt{n}}\mathbf{Z}'\hat{\mathbf{e}} \xrightarrow{d} (\mathbf{I}_\ell - \mathbf{A}) \cdot \frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}$$

- The matrix $\mathbf{I}_\ell - \mathbf{A}$ is an idempotent projection of rank $\ell - k$ – it projects out the k directions spanned by the columns of $\mathbf{Q}_{ZX}$
- Since $\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e} \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{Q}_{ZZ})$ and $(\mathbf{I}_\ell - \mathbf{A})$ is idempotent of rank $\ell - k$:

$$\frac{1}{\sigma^2}\left(\frac{1}{\sqrt{n}}\mathbf{Z}'\hat{\mathbf{e}}\right)'\mathbf{Q}_{ZZ}^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{Z}'\hat{\mathbf{e}}\right) \xrightarrow{d} \chi^2_{\ell-k}$$

- Substituting $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and $\frac{1}{n}\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \mathbf{Q}_{ZZ}$:

$$S = \frac{\hat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{\ell-k} \qquad \blacksquare$$

- **Key takeaway**: the $\ell - k$ degrees of freedom arise because estimation of β consumes exactly k of the ℓ moment conditions – the test is powered by what is left over

Intuition Behind the Sargan Statistic

Recall that the null hypothesis is:

$$H_0 : \mathbb{E}[\mathbf{Z}e] = 0$$

We cannot observe e, so we check the sample analogue – how far is $\frac{1}{n}\mathbf{Z}'\hat{\mathbf{e}}$ from zero?

If H_0 is true, $\frac{1}{n}\mathbf{Z}'\hat{\mathbf{e}}$ should be close to zero. If H_0 is false, it should be far from zero.

But "far from zero" for a vector needs to be made precise. We need a scalar measure of distance. The natural choice is a quadratic form:

$$\left(\frac{1}{n}\mathbf{Z}'\hat{\mathbf{e}}\right)' \mathbf{W} \left(\frac{1}{n}\mathbf{Z}'\hat{\mathbf{e}}\right)$$

for some weighting matrix $\mathbf{W}$. The question is what $\mathbf{W}$ to use.

The optimal choice is the inverse of the variance of $\frac{1}{\sqrt{n}}\mathbf{Z}'\hat{\mathbf{e}}$. Under homoskedasticity that variance is $\sigma^2\mathbf{Q}_{ZZ}$, so the natural weight is:

$$\mathbf{W} = (\sigma^2\mathbf{Q}_{ZZ})^{-1}$$

Plugging in sample analogues gives exactly S:

$$S = \frac{\hat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}}{\hat{\sigma}^2}$$

This is precisely a **standardized distance** – a **Mahalanobis-type** measure of how far the sample moment vector $\mathbf{Z}'\hat{\mathbf{e}}$ is from zero, scaled by its own variance. Large S means the instruments are correlated with the residuals, which is evidence against H_0.