Advanced Econometrics I
EMET4314/8014
Semester 1, 2026

Juergen Meinecke
Research School of Economics
ANU

Problem Set 7
(due: Monday week 9, 11:00am)

Submission Instructions: Same as last week.

Exercises

Provide transparent derivations. Justify steps that are not obvious. Use self sufficient proofs. Make reasonable assumptions where necessary.

The linear model under endogeneity is

$$Y = X\beta + e$$
$$X = Z\pi + v$$

where $\mathrm{E}(e_i X_i) \neq 0$ and $\mathrm{E}(e_i Z_i) = 0$. Notice $\dim X = N \times K$, $\dim \beta = K \times 1$, $\dim Z = N \times L$, $\dim \pi = L \times K$, and $\dim v = N \times K$.
The source of the endogeneity is correlation between the two error terms, write

$$e = v\rho + w$$

where $\mathrm{E}(v_i w_i) = 0$. Notice $\dim \rho = K \times 1$, and $\dim w = N \times 1$.
Combining, we obtain

$$Y = X\beta + v\rho + w \tag{1}$$

(a) You have available a random sample (X_i, Y_i, v_i). You are running a regression of Y on X and v. Using linear algebra, define the OLS estimator of β in equation (1). Call it $\hat{\beta}_0^{\text{OLS}}$.
(Hint: Use the *partitioned regression* result on the next page.)

(b) Prove that $\hat{\beta}_0^{\text{OLS}} = \beta + o_p(1)$.

(c) You do NOT have available a random sample (X_i, Y_i, v_i). Instead, you have available a random sample (X_i, Y_i, Z_i). You cannot run a regression of Y on X and v, but you can instead run a regression of Y on X and $\hat{v}$ where $\hat{v}$ is the first stage residual.

Using $\hat{v}$ in place of v in equation (1), define the OLS estimator of β using linear algebra. Call it $\hat{\beta}_1^{\text{OLS}}$.

Prove or disprove: $\hat{\beta}_1^{\text{OLS}} = (X'P_Z X)^{-1} X' P_Z Y$.

(d) Which estimator do you prefer: $\hat{\beta}_0^{\text{OLS}}$ or $\hat{\beta}_1^{\text{OLS}}$? No need to prove anything here, just give a quick intuitive statement.

Partitioned Regression and Frisch-Waugh-Lovell Theorem

Partition the linear regression model like so:

$$\begin{aligned} Y &= X\beta + e \\ &= X_1\beta_1 + X_2\beta_2 + e \end{aligned}$$

where X_1 is of dimension $N \times K_1$ and X_2 is of dimension $N \times K_2$ with $K_1 + K_2 = K$ and $X = [X_1 \;\; X_2]$. Then how could you estimate β_1? Write down the normal equations

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat\beta_1^{\text{OLS}} \\ \hat\beta_2^{\text{OLS}} \end{bmatrix} = \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

Solving first for $\hat\beta_2^{\text{OLS}}$

$$\begin{aligned} \hat\beta_2^{\text{OLS}} &= (X_2'X_2)^{-1}X_2'Y - (X_2'X_2)^{-1}X_2'X_1\hat\beta_1^{\text{OLS}} \\ &= (X_2'X_2)^{-1}X_2'(Y - X_1\hat\beta_1^{\text{OLS}}) \end{aligned}$$

Similarly

$$\hat\beta_1^{\text{OLS}} = (X_1'X_1)^{-1}X_1'(Y - X_2\hat\beta_2^{\text{OLS}})$$

This has an interesting interpretation:
The OLS estimator $\hat\beta_2^{\text{OLS}}$ results from regressing Y on X_2 *adjusted for* $X_1\hat\beta_1^{\text{OLS}}$. This *adjustment* is crucial, obviously it wouldn't be quite right to claim that $\hat\beta_2^{\text{OLS}}$ results from regressing X_2 on Y only. That would only be true of $X_2'X_1 = 0$ which means that the sample covariance between the two sets of regressors is zero. Now, doing the math by plugging $\hat\beta_2^{\text{OLS}}$ into $\hat\beta_1^{\text{OLS}}$ and
letting $P_2 := X_2(X_2'X_2)^{-1}X_2'$ and $M_2 = I - P_2$:

$$\begin{aligned} \hat\beta_1^{\text{OLS}} &= (X_1'X_1)^{-1}X_1'Y - \cdots \\ &\quad (X_1'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'Y + \cdots \\ &\quad (X_1'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'X_1\hat\beta_1^{\text{OLS}} \\ &= (X_1'X_1)^{-1}X_1'Y - (X_1'X_1)^{-1}X_1'P_2Y + \cdots \\ &\quad (X_1'X_1)^{-1}X_1'P_2X_1\hat\beta_1^{\text{OLS}} \\ &= (X_1'X_1)^{-1}X_1'M_2Y + (X_1'X_1)^{-1}X_1'P_2X_1\hat\beta_1^{\text{OLS}} \end{aligned}$$

Multiplying both sides by $X_1'X_1$ and moving terms

$$\begin{aligned} X_1'M_2Y &= (X_1'X_1)\hat\beta_1^{\text{OLS}} - X_1'P_2X_1\hat\beta_1^{\text{OLS}} \\ &= (X_1'M_2X_1)\hat\beta_1^{\text{OLS}} \end{aligned}$$

The end result (and also symmetrically for $\hat\beta_2^{\text{OLS}}$):

$$\begin{aligned} \hat\beta_1^{\text{OLS}} &= (X_1'M_2X_1)^{-1}X_1'M_2Y \\ \hat\beta_2^{\text{OLS}} &= (X_2'M_1X_2)^{-1}X_2'M_1Y \end{aligned}$$

Remember that M_1 and M_2 are *residual maker* matrices:

$$
\begin{aligned}
M_2X_1 &=: \widetilde{X}_1 && \text{is the residual in the regression of } X_1 \text{ on } X_2\\
M_2Y &=: \widetilde{Y}_2 && \text{is the residual in the regression of } Y \text{ on } X_2
\end{aligned}
$$

At the same time M_1 and M_2 are symmetric and idempotent
(that is $M_1 = M_1' = M_1M_1$)

$$
\begin{aligned}
\hat{\beta}_1^{\text{OLS}} &= ((M_2X_1)'(M_2X_1))^{-1}\,((M_2X_1)'(M_2Y)) = (\widetilde{X}_1'\widetilde{X}_1)^{-1}\widetilde{X}_1'\widetilde{Y}_2\\
\hat{\beta}_2^{\text{OLS}} &= ((M_1X_2)'(M_1X_2))^{-1}\,((M_1X_2)'(M_1Y)) = (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{Y}_1
\end{aligned}
$$

There's a lot of intuition included here. This harks back all the way to Gram Schmidt orthogonalization. To obtain $\hat{\beta}_1^{\text{OLS}}$, you regress a *version* of Y on a *version* of X_1.
The variable $\widetilde{X}_1$ is the version of X_1 from which the variation of X_2 has been removed. Likewise for $\widetilde{Y}_2$. Put differently, they are the versions of Y and X_1 in which the influence of X_2 has been *partialled out* or *netted out*.