

**Study Group Questions # 8**

In this study group assignment, you explore how the sampling properties of the Instrumental Variables estimators depend on the “strength” of the instruments. Our equation of interest is the following linear regression model,

$$y_i = x_i\beta_0 + u_i, \quad i = 1, 2, \dots, N, \quad (1)$$

where  $x_i$  is a scalar variable that is generated via the “first-stage” regression equation,

$$x_i = z'_i\gamma_0 + w_i, \quad (2)$$

where  $z_i$  is a  $q \times 1$  vector for some  $q > 2$  and  $\gamma_0$  is the vector of finite coefficients. It is convenient to set  $v_i = (u_i, w_i)'$ . Assume the following conditions hold.

**Assumption 1** (i)  $\{z_i\}_{i=1}^N$  is fixed in repeated samples with  $\|z_i\| < \infty$  for all  $i$  and  $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N z_i z'_i = Q_{zz}$ , a finite, pd matrix of constants; (ii)  $\{v_i\}_{i=1}^N$  is a sequence of i.i.d. random vectors with a normal distribution with mean equal to  $0_2$  and variance-covariance matrix equal to  $\Sigma$ , with

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{u,w} \\ \sigma_{w,u} & \sigma_w^2 \end{bmatrix}.$$

While our primary focus is on the properties of the 2SLS estimator, part of the analysis uses the OLS estimator as a comparator. To present the formulae for the two estimators in the context of our model above, we introduce the following notation. Let  $y$ ,  $x$ ,  $u$  and  $w$  to be  $N \times 1$  vectors whose  $i^{th}$  elements are respectively  $y_i$ ,  $x_i$ ,  $u_i$  and  $w_i$ ; let  $Z$  be the  $N \times q$  matrix with  $i^{th}$  row  $z'_i$ , and  $P_z = Z(Z'Z)^{-1}Z'$ . The generic formula for the OLS and 2SLS estimator from lectures specialize in the context of our model here to the following:

$$\hat{\beta}_{OLS} = \frac{x'y}{x'x}, \quad (3)$$

$$\hat{\beta}_{2SLS} = \frac{x'P_z y}{x'P_z x}. \quad (4)$$

As discussed in the Tutorial session, the approximate bias of the OLS and 2SLS estimators is given by the following expressions:

$$bias(\hat{\beta}_{OLS}) \approx \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \left( \frac{1}{\mu/N + 1} \right), \quad (5)$$

$$bias(\hat{\beta}_{2SLS}) \approx \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \left( \frac{q - 2}{\mu} \right), \quad (6)$$

where  $\mu$  is the *concentration parameter* discussed in the Tutorial session that is,

$$\mu = \frac{\gamma_0' Z' Z \gamma_0}{\sigma_w^2}.$$

Please enter the details of the group:

Group name	Group Matrix
Student ID 1	11335127
Student ID 2	10710007
Student ID 3	10704589
Student ID 4	11465531
Student ID 5	

1. Show that  $\mu \rightarrow \infty$  and  $\mu/N \rightarrow c$  as  $N \rightarrow \infty$  where  $c$  is a finite positive constant that you must specify as part of your answer. Note: you may quote results established in the Tutorial session but explain clearly how you apply them.

Examine  $\mu = \frac{\gamma_0^T z^T z \gamma_0}{\delta w}$  with  $z$  is a  $N \times q$  matrix and  $\gamma_0$  is  $q \times 1$ .

We can rewrite  $\mu = \frac{\gamma_0^T z^T}{\delta w} \times \frac{z \gamma_0}{\delta w} = b^T b$  with  $b = \frac{z \gamma_0}{\delta w}$  is a  $N \times 1$  matrix,

$$\Rightarrow \mu = \sum_{i=1}^N b_i^2 \text{ with } b_i \text{ is the } i^{\text{th}} \text{ row of matrix } b; \mu \text{ is non-negative.}$$

(if  $\gamma_0 \neq 0$ )

Therefore, it follows that as  $N \rightarrow \infty$ ;  $\sum_{i=1}^N b_i^2 \rightarrow \infty$  or  $\mu \rightarrow \infty$  when  $\gamma_0 \neq 0$ .

Examine  $\frac{\mu}{N} = \frac{\gamma_0^T z^T z \gamma_0}{\delta w N} = \frac{\gamma_0^T}{\delta w} \times \frac{z^T z}{N} \times \frac{\gamma_0}{\delta w} = A^T \frac{z^T z}{N} A$ ;  $A = \frac{\gamma_0}{\delta w}$ , a  $q \times 1$  matrix of constant.

Then,  $\lim_{N \rightarrow \infty} \frac{\mu}{N} = A^T Q_{zz} A$  because  $z^T z = \sum_{i=1}^N z_i z_i^T$  and  $\lim_{N \rightarrow \infty} \sum_{i=1}^N z_i z_i^T = Q_{zz}$

We can rewrite the quadratic form  $A^T Q_{zz} A$  as  $\sum_{j=1}^q \sum_{i=1}^q Q_{ij} A_i A_j$  in which:

$$\left\{ \begin{array}{l} Q_{ij} \text{ is the } i^{\text{th}} - j^{\text{th}} \text{ element of } Q_{zz} \\ A_i \text{ is the } i^{\text{th}} \text{ element of } A = \frac{\gamma_0}{\delta w} \\ A_j \text{ is the } j^{\text{th}} \text{ element of } A. \end{array} \right.$$

$\Rightarrow \lim_{N \rightarrow \infty} \frac{\mu}{N} = \sum_{j=1}^q \sum_{i=1}^q Q_{ij} A_i A_j$  which is a non-negative constant.

(it's positive if  $\gamma_0 \neq 0$ ).

2. Under what condition is the OLS estimator approximately unbiased that is, its approximate bias is zero. Interpret this condition in the context of the model described above.

We have  $\text{bias}(\hat{\beta}_{OLS}) \approx \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \left( \frac{1}{n/N + 1} \right)$

From Question 1, we know that  $n/N$  converges to a finite non-negative constant. Therefore,  $\text{bias}(\hat{\beta}_{OLS}) = 0 \Leftrightarrow \sigma_{u,w} = 0$ , which means  $u$  and  $w$  are uncorrelated.

In our setting,  $u$  is the error term from the structural equation (1) and  $w$  is the error term from the first stage regression (2). Combined with the notion from Tutorial Question 1 that  $E[x_i u_i] = \sigma_{u,w}$  we can interpret that the OLS estimator is approximately unbiased if  $x_i$  is exogenous.

3. Under what conditions is the 2SLS estimator approximately unbiased. Interpret these conditions in the context of the model described above.

We have  $\text{bias}(\hat{\beta}_{2SLS}) \approx \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \left( \frac{q-2}{n} \right)$  with  $q > 2$ .

There are three cases for which  $\text{bias}(\hat{\beta}_{2SLS}) = 0$

Case 1:  $N \rightarrow \infty$  and  $\sigma_{u,w} \neq 0$

As  $N \rightarrow \infty$ ,  $n \rightarrow \infty$  (Question 1)

$$\lim_{N \rightarrow \infty} \text{bias}(\hat{\beta}_{2SLS}) = \lim_{N \rightarrow \infty} \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \left( \frac{q-2}{n} \right) = \left( \frac{\sigma_{u,w}}{\sigma_w^2} \right) \times 0 = 0$$

Case 2:  $N$  is not sufficiently large and  $\sigma_{u,w} = 0$ .

Then, similar to Question 2,  $\text{bias}(\hat{\beta}_{2SLS}) \approx 0$ .

Case 3: Both  $N$  is large and  $\sigma_{u,w} = 0$  can exist at the same time.

In our setting, the conditions under which the 2SLS estimator is approximately unbiased includes:

- 1)  $x_i$  is endogenous but the sample size is very large, toward  $\infty$ .
- 2)  $x_i$  is exogenous in small samples.
- 3)  $x_i$  is exogenous and the sample size is large.

4. Define the relative approximate bias of 2SLS to OLS as:

$$rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS}) = \frac{bias(\hat{\beta}_{2SLS})}{bias(\hat{\beta}_{OLS})}.$$

What is the differential of  $rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS})$  with respect to a change in  $\mu$  holding all else constant? Interpret your result.

$$\text{Braniac } rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS}) = \frac{bias(\hat{\beta}_{2SLS})}{bias(\hat{\beta}_{OLS})} = \frac{\left(\frac{\sigma_{\epsilon, w}}{\sigma_w^2}\right)\left(\frac{q-2}{N}\right)}{\left(\frac{\sigma_{\epsilon, w}}{\sigma_w^2}\right)\left(\frac{1}{N/2+1}\right)}$$

$$\Rightarrow rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS}) = \left(\frac{q-2}{N}\right)\left(\frac{1}{N} + 1\right) = \frac{q-2}{N} + \frac{q-2}{N} > 0$$

$$\Rightarrow d(rbias) = \frac{2-q}{N^2} d\mu \text{ is the differential of rbias with respect to } \mu.$$

$\Rightarrow$  Holding all else equal, a change in  $\mu$  associates with  $\frac{2-q}{N^2}$  change in rbias.

Since  $q > 2 \Rightarrow \frac{2-q}{N^2} < 0$ . Thus,  $\mu$  and rbias move in opposite directions.

We also have that  $\mu = \frac{\gamma_0 z' z \gamma_0}{\sigma_w^2}$  meaning  $\mu$  is increasing with  $|\gamma_0|$ , keeping all else unchanged

As  $\mu$  is the non-centrality parameter related to F-statistics to test  $H_0: \gamma_0 = 0$  vs.  $H_1: \gamma_0 \neq 0$ . When  $|\gamma_0|$  is substantially large, under  $H_1$ ,  $\mu$  is also large.

Moreover, when the sample size  $N \rightarrow \infty$ ,  $\mu \rightarrow \infty$  as well if  $\gamma_0 \neq 0$ .

Therefore, we can interpret that the relative approximate bias is reduced when  $z_i'$  is highly correlated with  $x_i$ . This effect will enhance when the sample size is large.

In conclusion, if the variation of  $x_i$  is well-explained by  $z_i$ , which means the relevance condition holds, relative bias between 2SLS and OLS estimator is reduced.

In this case, rbias will tend towards 0 if  $N$  goes towards  $\infty$ .

5. What happens to  $rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS})$  as  $\mu$  becomes close to zero holding all else constant?

We have  $rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS}) = \frac{q-2}{N} + \frac{q-2}{\mu}$

As  $\mu \rightarrow 0 \Rightarrow \frac{q-2}{\mu} \rightarrow \infty \Rightarrow rbias \rightarrow \infty$ .

Under  $H_0: \rho_0 = 0$ ,  $\mu$  tends towards 0 (even when the sample size is large). Which means that if  $z_i'$  is not relevant to  $x_i$ ,  $rbias(\hat{\beta}_{2SLS}, \hat{\beta}_{OLS})$  tends to be large. Even in case of weak instrument where  $z_i'$  is weakly relevant to  $x_i$ ,  $\mu$  is small making  $rbias$  substantially large.

Therefore, holding all else equal, if the variance in  $x_i$  is not well-explained by  $z_i'$ , relative approximate bias between OLS and 2SLS estimator is large.

6. Use your answers to Questions 2-5 to evaluate the following statement: "In a linear regression model with an endogenous regressor it is always preferable to base inferences on an IV estimator than to base inferences on the OLS estimator."

In the presence of endogeneity, we know (from Question 2) that OLS is not a good estimator since the approximate bias is not zero.

On the other hand, 2SLS estimator may be able to reduce the bias in large samples (Question 3). This is because when the correlation between  $Z$  and  $X$  significantly differs 0, the non-centrality parameter,  $\mu$ , will grow as sample size grows, leading to smaller bias. In that case, when  $Z$  well-explains  $X$  and the sample size is large, we can expect the relative bias between 2SLS estimator and OLS estimator to tend towards 0 (Question 4) and the 2SLS estimator is less biased than the OLS estimator. However, when  $Z$  and  $X$  are not strongly relevant, holding all else equal, the relative bias will become large (Question 5).

Therefore, only when we have  $Z$  that is strongly relevant to  $X$  and a large sample size that the use of 2SLS is preferable to OLS in case of endogeneity given all else unchanged.