Advanced Econometrics II
TA Session Problems No. 1

Agnieszka Borowska

06.01.2015

Note: this is only a draft of the problems discussed on Tuesday and might contain some typos or more or less imprecise statements. If you find some, please let me know.

1 Asymptotic Efficiency

1.1 Brief Recap

Model:

$$y = X\beta_0 + u,$$
$$\mathbb{E}[u^T u] = \sigma_0^2 \mathbb{I},$$
$$\text{plim}\left[n^{-1} X^T u\right] \neq 0.$$

Instruments: available $l \geq k$, i.e. a matrix W with dimension $n \times l$ (**overidentification**). We need to have k instruments, given by the matrix $\tilde{W}$, such that

$$\text{plim}\left[n^{-1}\tilde{W}\tilde{W}^T u\right] = 0, \qquad \text{(validity)}$$
$$\text{plim}\left[n^{-1}\tilde{W}^T X\right] = S_{\tilde{W}X}, \qquad \text{(relevance)}$$
$$\text{plim}\left[n^{-1}\tilde{W}\right] = S_{\tilde{W}\tilde{W}}, \qquad \text{(technicality)}$$

with $\text{rank}(S_{\tilde{W}X}) = k = \text{rank}(S_{\tilde{W}\tilde{W}})$ (finite, full rank).

1.2 Optimal instruments

How to get $\tilde{W}$? Use a selection matrix J, such that

$$\tilde{W} = WJ.$$

to minimize the asymptotic variance of

$$\hat{\beta}_{IV} = \left((WJ)^T X\right)^{-1}(WJ)^T y$$
$$= \beta_0 + \left((WJ)^T X\right)^{-1}(WJ)^T u.$$

This gives us the following **convenient expression** for deriving asymptotic distribution

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta_0\right) = \underbrace{\left(\frac{1}{n}(WJ)^T X\right)^{-1}}_{(\star)} \underbrace{\left(\frac{1}{\sqrt{n}}(WJ)^T u\right)}_{(\star\star)}.$$

We have

$$(\star) \xrightarrow{p} S_{WX}^{-1}, \qquad \text{(Slutsky)},$$
$$(\star\star) \xrightarrow{d} \mathcal{N}\left(0, \sigma_0^2 S_{WJWJ}\right), \qquad \text{(CLT \& MW)},$$

where

$$S_{WJWJ} = \text{plim}\left[n^{-1}(WJ)^T WJ\right],$$
$$S_{WJX} = \text{plim}\left[n^{-1}(WJ)^T X\right]$$

Recall the **Cremér's Theorem**, which deals with a linear transformation of a normal variate z_n. If

$$\sqrt{n}\left(z_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right), \qquad (K \times 1)$$
$$H_n \xrightarrow{p} H, \qquad (K \times J),$$

then

$$\sqrt{n} H_n^T \left(z_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, H^T \Sigma H\right).$$

Then, by this theorem

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta_0\right) \xrightarrow{d} \mathcal{N}\Big(0, \underbrace{\sigma_0^2 S_{WJX}^{-1} S_{WJWJ}\left(S_{WJX}^{-1}\right)^T}_{\text{AVar}}\Big).$$

Hence, we have obtained the asymptotic covariance matrix in a **sandwich form**:

$$\text{AVar}(WJ) = \sigma_0^2 S_{WJX}^{-1} S_{WJWJ}\left(S_{WJX}^{-1}\right)^T \tag{1}$$
$$= \sigma_0^2 \text{plim}\left[n^{-1} X^T P_{WJ} X\right]^{-1}, \tag{2}$$

with the **precision matrix** corresponding to (2)

$$\sigma_0^2 \text{plim}\left[n^{-1} X^T P_{WJ} X\right],$$

To get the **optimal** J, we need to eliminate the sandwich in (1). And to achieve this, we need

$$S_{WJWJ} = S_{WJX},$$
$$\text{plim}\left[n^{-1}(WJ)^T WJ\right] = \text{plim}\left[n^{-1} X^T WJ\right],$$
$$(WJ)^T WJ = X^T WJ,$$
$$J^T W^T W = X^T W,$$
$$J^T = X^T W (W^T W)^{-1},$$
$$J = \underbrace{(W^T W)^{-1} W^T X}_{\text{Optimal choice}}. \tag{3}$$

Notice, that under this optimal choice (3)

$$WJ = W(W^T W)^{-1} W^T X$$
$$= P_W X,$$

so that the required instrument selection $\tilde{W} = WJ$ is just a projection of X on the span of W. Then, the asymptotic variance takes the form

$$\text{AVar}(P_W X) = \sigma_0^2 \left(\text{plim}\left[n^{-1} X^T P_{P_W X} X\right]\right)^{-1}. \tag{4}$$

Recall, however, that

$$P_{P_W X} = P_W X \left((P_W X)^T P_W X\right)^{-1} (P_W X)^T$$
$$= P_W,$$

so that (4) becomes

$$\text{AVar}(W) = \sigma_0^2 \left(\text{plim}\left[n^{-1} X^T P_W X\right]\right)^{-1}. \tag{5}$$

1.3 Asymptotic efficiency

Summing up: with an arbitrary selection matrix J the asymptotic covariance matrix is given by (2), while with the optimal selection matrix, equal to (3), it is given by (5). To show asymptotic efficiency of the IV estimator derived with the optimally selected instruments, we need to show that the difference between the asymptotic variance matrices of the non-optimal and of the optimal estimator, i.e.

$$\text{AVar}(WJ) - \text{AVar}(P_W X) > 0,$$

is **positive semidefinite** (PSD).

From Exercise 3.8 we know that the matrix

$$\left(X^T P_{WJ} X\right)^{-1} - \left(X^T P_W X\right)^{-1}$$

is positive definite (PD) iff the matrix

$$X^T P_W X - X^T P_{WJ} X$$

is PD, which holds true also for the PSD case. Hence, to show the required result we can work with the expression without matrix inverses

$$X^T P_W X - X^T P_{WJ} X,$$

which is proportional to the difference between the two precision matrices (with plim's and n's dropped).

Since J is a *selection* matrix, $\text{lin}(WJ) \subset \text{lin}(W)$, so that P_W is a matrix of a projection on a bigger subspace then in the case of P_{WJ}. Hence,

$$P_{WJ} = P_{WJ} P_W = P_W P_{WJ},$$

so that

$$\begin{aligned} X^T P_W X - X^T P_{WJ} X &= X^T \left(P_W - P_{WJ}\right) X \\ &= X^T \left(P_W - P_W WJ P_W\right) X \\ &= X^T P_W \underbrace{\left(\mathbb{I} - P_{WJ}\right)}_{M_{WJ}} P_W X \\ &= X^T P_W M_{WJ} P_W X \end{aligned}$$

where M_{WJ} is an orthogonal projection matrix. This means that, indeed,

$$X^T P_W X - X^T P_{WJ} X$$

is PSD, which implies that

$$\text{AVar}(WJ) - \text{AVar}(W)$$

is a **positive semidefinite matrix**.

Conclusion: $\tilde{W} = P_W X$ is the optimal choice of instrumental variables by the criterion of asymptotic variance, leading o an asymptotically efficient GIV estimator.

However, an **efficiency gain** is potentially available if the space spanned by the columns of W is made larger, i.e. by adding extra instruments. As we will show below, appending new columns to W will reduce the asymptotic covariance matrix.

1.4 Ex. DM 8.8

Suppose that W_1 and W_2 are, respectively, $n \times l_1$ and $n \times l_2$ matrices of instruments, and that W_2 consists of W_1 plus $l_2 - l_1$ additional columns. Prove that the generalized IV estimator using W_2 is asymptotically more efficient than the generalized IV estimator using W_1. To do this, you need to show that the matrix

$$\left(X^T P_{W_1} X\right)^{-1} - \left(X^T P_{W_2} X\right)^{-1}$$

is PSD. Hint: see Exercise 3.8.

Exactly as we have done above, we use Exercise 3.8 to argue that the matrix $\left(X^T P_{W_1} X\right)^{-1} - \left(X^T P_{W_2} X\right)^{-1}$ is PSD iff the matrix $X^T P_{W_2} X - X^T P_{W_1} X$ is PSD, and we work with the expression without matrix inverses

$$X^T P_{W_2} X - X^T P_{W_1} X. \tag{6}$$

Notice that $\text{lin}(W_1) \subset \text{lin}(W_2)$, because

$$X_2 = [X_1, X_*],$$

where W_* is some set of instruments, different than W_1, of dimension $n \times (l_2 - l_1)$. Thus,

$$P_{W_1} P_{W_2} = P_{W_2} P_{W_1} = P_{W_1}.$$

Hence, (6) can be rewritten as follows

$$\begin{aligned} X^T P_{W_2} X - X^T P_{W_1} X &= X^T \left(P_{W_2} - P_{W_1}\right) X \\ &= X^T \left(P_{W_2} - P_{W_2} P_{W_1} P_{W_2}\right) X \\ &= X^T P_{W_2} \underbrace{\left(\mathbb{I} - P_{W_1}\right)}_{M_{W_1}} P_{W_2} X \\ &= X^T P_{W_2} M_{W_1} P_{W_2} X, \end{aligned} \tag{7}$$

which is PSD since M_{W_1} is an orthogonal projection matrix. To finish the exercise, divide (7) by n and let n go to infinity.

Conclusion: including of more instruments increases the efficiency.

2 Identifiability and Consistency

2.1 Criterion function

$$\begin{aligned} \text{OLS: } Q(\beta, y) &= (y - X\beta)^T (y - X\beta), \\ \text{GIV: } Q(\beta, y) &= (y - X\beta)^T P_W (y - X\beta). \end{aligned}$$

The former is simply the SSR. The latter can be expressed as

$$Q(\beta, y) = y^T P_W y + \beta^T X^T P_W X \beta - 2\beta^T X^T P_W y,$$

which minimised wrt β yields the following FOC

$$2X^T P_W X\beta - 2X^T P_W y = 0.$$

The solution to this problem is

$$\hat{\beta}_{IV} = \left(X^T P_W X\right)^{-1} X^T P_W y, \tag{8.29}$$

the **generalized IV estimator**, which is a generalisation of the simple IV estimator

$$\hat{\beta}_{IV} = \left(W^T X\right)^{-1} W^T y, \tag{8.12}$$

2.2 Identification

Identification condition:

$$\text{plim}\left[n^{-1} Q(\beta, y)\right] \begin{cases} = 0, & \text{if } \beta = \beta_0, \\ > 0, & \text{if } \beta \neq \beta_0, \end{cases}$$

i.e. the plim the IV criterion function divided by n has a **unique global minimum** at $\beta = \beta_0$[1].

[1] Cf. Exercise DM 8.5.

The **asymptotic identification condition:**

- the simple IV estimator

$$S_{WX} \equiv \text{plim}\left[n^{-1}W^T X\right] \quad \text{is deterministic and non-singular,} \tag{8.14}$$

- the general IV estimator

$$S_{XW}\left(S_{WW}\right)^{-1} S_{WX} \quad \text{has a full rank.} \tag{8}$$

2.3 Consistency

The **consistency condition:**

$$\text{plim}\left[n^{-1}W^T u\right] = 0, \tag{8.16}$$

Under the key assumption (8.16), the two asymptotic identification conditions are sufficient for consistency.

2.4 Ex. DM 8.6

Under assumption (8.16) and the asymptotic identification condition that $S_{XW}\left(S_{WW}\right)^{-1} S_{WX}$ has full rank, show that the GIV estimator $\hat{\beta}_{IV}$ is consistent by explicitly computing the probability limit of the estimator for a DGP such that $y = X\beta_0 + u$.

The GIV estimator is given by

$$\begin{aligned}
\hat{\beta}_{IV} &= \left(X^T P_W X\right)^{-1} X^T P_W y \\
&= \left(X^T P_W X\right)^{-1} X^T P_W \left(X\beta_0 + u\right) \\
&= \left(X^T W \left(W^T W\right)^{-1} W^T X\right)^{-1} X^T W \left(W^T W\right)^{-1} W^T \left(X\beta_0 + u\right) \\
&= \beta_0 + \underbrace{\left(X^T W \left(W^T W\right)^{-1} W^T X\right)^{-1} X^T W \left(W^T W\right)^{-1} W^T u}_{(*)},
\end{aligned}$$

since $P_W = W\left(W^T W\right)^{-1} W^T$.

Consider $(*)$. Its plim stays the same if we divide by n all the subsequent "blocks" i.e. $X^T W$, $W^T W$, $W^T W$ and $W^T u$. Then, each factor has a deterministic plim and the plim of $\hat{\beta}_{IV}$ becomes

$$\begin{aligned}
\text{plim}\left[\hat{\beta}_{IV}\right] &= \text{plim}\left[\beta_0\right] + \text{plim}\left[(*)\right] \\
&= \beta_0 + \underbrace{\left(S_{XW}\left(S_{WW}\right)^{-1} S_{WX}\right)}_{(**)}{}^{-1} S_{XW}\left(S_{WW}\right)^{-1} \cdot \underbrace{\text{plim}\left[n^{-1}W^T u\right]}_{(***)}.
\end{aligned}$$

Condition (8) guarantees that $(**)$ is non-singular, while assumption (8.16) that $(***)$ is a zero vector. Therefore, these two conditions indeed lead to

$$\text{plim}\left[\hat{\beta}_{IV}\right] = \beta_0,$$

which means that the GIV estimator consistent.