Advanced Econometrics II
TA Session Problems No. 3

Agnieszka Borowska

20.01.2015

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.

Model

$$\begin{aligned} y &= X\beta + u, \\ \mathbb{E}(uu^T) &= \Omega, \\ \mathbb{E}(u_t|X_t) &\neq 0, \\ \mathbb{E}(u_t|W_t) &= 0, \\ W_t &\in \Omega_t. \end{aligned} \tag{1}$$

The last assumption says the instruments are predetermined. Moreover, we have $l > k$, i.e. overidentification.

1. Efficient GMM
   select J for $Z \equiv WJ$ (optimal choice of the selection matrix, given the instruments)
2. Fully efficient GMM
   select W (the best choice of instruments W out of all the possible valid instruments, given the information set Ω_t); in GLS spirit.

Efficient GMM

Aim: given instruments W, find the optimal selection matrix J in the case when Ω not proportional to the identity matrix, i.e. when $\Omega \neq \sigma^2\mathbb{I}$.

When we use WJ as instruments, the **moment conditions** are

$$J^T W^T (y - X\beta) = 0,$$

so that the asymptotic distribution of the estimator $\hat{\beta}$ which solves them follows from

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) = \left(\frac{1}{n} J^T W^T X\right)^{-1} \left(\frac{1}{\sqrt{n}} J^T W^T u\right) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{AVar}(WJ)\big),$$

where the asymptotic covariance matrix is given by

$$\mathrm{AVar}(WJ) = \operatorname{plim} \left(\frac{1}{n} X^T WJ\big(J^T W^T \Omega WJ\big)^{-1} J^T W^T X\right)^{-1}.$$

The sandwich eliminating choice of J:

$$J^* = (W^T\Omega W)^{-1} W^T X,$$

so that the asymptotic covariance matrix becomes

$$\mathrm{AVar}(WJ^*) = \operatorname{plim} \left(\frac{1}{n} X^T W\big(W^T \Omega W\big)^{-1} W^T X\right)^{-1}. \tag{2}$$

The difference between $\mathrm{AVar}(WJ)$ and $\mathrm{AVar}(WJ^*)$ is PSD, hence, indeed, J^* is the optimal choice. The resulting **efficient GMM estimator** has the form

$$\hat{\beta}_{GMM} = \left(X^T W\big(W^T\Omega W\big)^{-1} W^T X\right)^{-1} X^T W\big(W^T\Omega W\big)^{-1} W^T y. \tag{3}$$

Fully efficient GMM

Aim: find the optimal choice of the instruments in the case when Ω not proportional to the identity matrix, i.e. when $\Omega \neq \sigma^2\mathbb{I}$.

1° First, suppose that X is **exogenous**, so we can use it as the instruments: $W = X$. Then, the efficient GMM estimator (3) boils down to the OLS estimator:

$$\begin{aligned}\hat{\beta}_{GMM} &= \left(X^T W \left(W^T\Omega W\right)^{-1} W^T X\right)^{-1} X^T W \left(W^T\Omega W\right)^{-1} W^T y,\\ &= \left(X^T X \left(X^T\Omega X\right)^{-1} X^T X\right)^{-1} X^T X \left(X^T\Omega X\right)^{-1} X^T y,\\ &= \left(X^T X\right)^{-1} X^T\Omega X \left(X^T X\right)^{-1} X^T X \left(X^T\Omega X\right)^{-1} X^T y,\\ &= \left(X^T X\right)^{-1} X^T\Omega X \left(X^T\Omega X\right)^{-1} X^T y,\\ &= \left(X^T X\right)^{-1} X^T y,\\ &= \hat{\beta}_{OLS}.\end{aligned}$$

However, in the case when $\Omega \neq \sigma^2\mathbb{I}$, the OLS estimator is **not efficient**, as the efficient one is the GLS estimator $\hat{\beta}_{GLS}$ given by[1]

$$\begin{aligned}\hat{\beta}_{GLS} &= \big(\underbrace{X^T\Omega^{-1}}_{W^T} X\big)^{-1} \underbrace{X^T\Omega^{-1}}_{W} y\\ &= \left(W^T X\right)^{-1} W^T y\\ &= \hat{\beta}_{IV},\end{aligned}$$

with $W = \Omega^{-1}X$. So when $\Omega \neq \sigma^2\mathbb{I}$, the optimal instruments are no longer given by

$$\mathbb{E}[X_t|\Omega_t] \equiv \bar{X}_t = X_t,$$

i.e. the predetermined part of the explanatory variables X, as they are equal to $\Omega^{-1}X$.

2° Next, suppose that some variables in X are **not predetermined**, so we need to instrument for them. Simple solution

$$\Omega^{-1}\bar{X}$$

does not work because even if $\bar{X}_t$ is predetermined, $\Omega^{-1}\bar{X}$ is not due to **serial correlation**.

Hence, **GLS approach** - the aim: construct Ψ, $n \times n$, such that $\Omega^{-1} = \Psi\Psi^T$. Then we can premultiply (1) by Ψ^T to get the **transformed model**

$$\Psi^T y = \Psi^T X\beta + \Psi^T u, \tag{4}$$

so that the covariance matrix of the transformed error vector $\Psi^T u$ is

$$\begin{aligned}\mathbb{E}\left[\Psi^T u u^T \Psi|\Omega_t\right] &= \mathbb{E}\left[\Psi^T \Omega \Psi|\Omega_t\right]\\ &= \mathbb{E}\left[\Psi^T (\Psi\Psi^T)^{-1} \Psi|\Omega_t\right]\\ &= \mathbb{I}_n,\end{aligned}$$

the identity matrix. Because of endogeneity we need to find Z, a matrix of instruments **for the transformed model** (4), such that the **theoretical moment conditions**

$$\mathbb{E}\left[Z^T\Psi^T(y - X\beta)\right] = 0, \tag{5}$$

are satisfied. Notice that for (5) to hold we need

$$\mathbb{E}\left[\left(\Psi^T u\right)_t |Z_t\right] = 0,$$

so the instruments are valid wrt to the transformed error terms.

[1] Cf. Section 7.2 in DM.

Notice, that in the case from 1° of exogenous X, when the optimal instruments for the untransformed model (1) were $\Omega^{-1}X$, the optimal choice of Z for the transformed model (4) is

$$Z = \Psi^T X,$$

as then

$$\begin{aligned} 0 &= \mathbb{E}\left[Z^T \Psi^T (y - X\beta)\right] \\ &= \mathbb{E}\left[X^T \Psi \Psi^T (y - X\beta)\right] \\ &= \mathbb{E}\left[X^T \Omega^{-1} (y - X\beta)\right], \end{aligned}$$

the same as the theoretical moment conditions for the exogenous case.

Usually it is possible to find Ψ such that the linear combination of u's, $(\Psi^T u)_t$, are **innovations** wrt Ω_t, i.e.

$$\mathbb{E}\left[\left(\Psi^T u\right)_t |\Omega_t\right] = 0.$$

When X is not exogenous and $\Omega \neq \sigma^2 \mathbb{I}$, we need to find $\bar{X}$ which are **implicitly defined** by

$$\mathbb{E}\left[\left(\Psi^T X\right)_t |\Omega_t\right] = (\Psi^T \bar{X})_t \tag{6}$$

so that $\Psi^T \bar{X}$ is are predetermined and we can use them as instruments Z. This is not an easy task and needs to be handled on a **case-by-case basis**.

So we claim that setting

$$Z = \Psi^T \bar{X}$$

with $\bar{X}$ implicitly defined in (6) is the **optimal choice** in our general setup. Let's check it. First, notice that this choice leads to (5) becoming

$$\begin{aligned} \mathbb{E}\left[Z^T \Psi^T (y - X\beta)\right] &= \mathbb{E}\left[\bar{X}^T \Psi \Psi^T (y - X\beta)\right] \\ &= \mathbb{E}\left[\bar{X}^T \Omega^{-1} (y - X\beta)\right] \\ &= 0, \end{aligned}$$

which result in the following **efficient GMM estimator**

$$\hat{\beta}_{EGMM} = \left(\bar{X}^T \Omega^{-1} \bar{X}\right)^{-1} \bar{X}^T \Omega^{-1} y.$$

Its **asymptotic covariance matrix** can be obtained by plugging into (2)

$$\begin{aligned} W &:= \Psi^T \bar{X}, \\ X &:= \Psi^T X, \\ \Omega &:= \mathbb{I}, \end{aligned}$$

(notice that we need to use **transformed error** terms) to obtain

$$\begin{aligned} \mathrm{AVar}(\hat{\beta}_{EGMM}) &= \mathrm{plim}\, \left(\frac{1}{n} X^T \Psi \Psi^T \bar{X} \left(\bar{X}^T \Psi \mathbb{I} \Psi^T \bar{X}\right)^{-1} \bar{X}^T \Psi \Psi^T X\right)^{-1} \\ &= \mathrm{plim}\, \left(\frac{1}{n} X^T \Omega^{-1} \bar{X} \left(\frac{1}{n} \bar{X}^T \Omega^{-1} \bar{X}\right)^{-1} \underbrace{\frac{1}{n} \bar{X}^T \Omega^{-1} X}_{(*)}\right)^{-1}. \end{aligned}$$

Next, consider $(*)$. Notice that for any M such that $M_t \in \Omega_t$ we have

$$\begin{aligned} \mathrm{plim}\, \frac{1}{n} M^T \Psi^T X &= \mathrm{plim}\, \frac{1}{n} \mathbb{E}\left[M^T \Psi^T X |\Omega_t\right] \\ &= \mathrm{plim}\, \frac{1}{n} \mathbb{E}\left[M^T \Psi^T \bar{X} |\Omega_t\right] \\ &= \mathrm{plim}\, \frac{1}{n} M \Psi^T \bar{X}. \end{aligned}$$

Since by (6) we have $\left(\Psi^T \bar{X}\right)_t \in \Omega_t$ so that

$$\begin{aligned}
\text{plim } \frac{1}{n} \bar{X} \Omega^{-1} X &= \text{plim } \frac{1}{n} \bar{X} \Psi \Psi^T X \\
&= \text{plim } \frac{1}{n} \bar{X} \Psi \Psi^T \bar{X} \\
&= \text{plim } \frac{1}{n} \bar{X} \Omega^{-1} \bar{X}.
\end{aligned}$$

This simplifies the asymptotic covariance matrix to

$$\text{AVar}(\hat{\beta}_{EGMM}) = \text{plim } \left(\frac{1}{n} \bar{X}^T \Omega^{-1} \bar{X} \right)^{-1}. \tag{7}$$

Is (7) "better" than (2)?
Suppose that we use $Z = WJ$, where W are some predetermined instruments ($W_t \in \Omega_t$). Then, the moment conditions are

$$Z^T \Psi^T (y - X\beta) = J^T W^T \Psi^T (y - X\beta) = 0,$$

which yield the following solution

$$\hat{\beta} = \left(J^T W^T \Psi^T X \right)^{-1} J^T W^T \Psi^T y.$$

Its asymptotic covariance matrix has the following sandwich form

$$\text{plim } \left(\frac{1}{n} X^T \Psi W J \left(J^T W^T W J \right)^{-1} J^T W^T \Psi^T X \right)^{-1}. \tag{8}$$

The sandwich can be eliminated when

$$W^T \Psi^T X = W^T W J,$$

which gives the optimal choice of J:

$$J^* = (W^T W)^{-1} W^T \Psi^T X.$$

Then

$$\begin{aligned}
J^T W^T \Psi^T X &= ((W^T W)^{-1} W^T \Psi^T X)^T W^T \Psi^T X \\
&= X^T \Psi \underbrace{W (W^T W)^{-1} W^T}_{P_W} \Psi^T X \\
&= X^T \Psi P_W \Psi^T X, \\
J^T W^T W J &= X^T \Psi W (W^T W)^{-1} W^T W (W^T W)^{-1} W^T \Psi^T X \\
&= X^T \Psi \underbrace{W (W^T W)^{-1} W^T}_{P_W} \Psi^T X \\
&= X^T \Psi P_W \Psi^T X.
\end{aligned}$$

So with J^*, the asymptotic covariance matrix (8) of $\hat{\beta}$ becomes

$$\text{plim } \left(\frac{1}{n} X^T \Psi P_W \Psi^T X \right)^{-1}.$$

Since we assumed that the instruments are predetermined, we can use the reasoning as above to obtain

$$\text{plim } \left(\frac{1}{n} X^T \Psi P_W \Psi^T X \right)^{-1} = \text{plim } \left(\frac{1}{n} \bar{X}^T \Psi P_W \Psi^T \bar{X} \right)^{-1},$$

The difference between the two precision matrices corresponding to (7) and (2) is then given by

$$\begin{aligned}
\bar{X}^T \Psi \Psi^T \bar{X} - \bar{X}^T \Psi P_W \Psi^T \bar{X} &= \bar{X}^T \Psi \left(\mathbb{I} - P_W \right) \Psi^T \bar{X} \\
&= \bar{X}^T \Psi M_W \Psi^T \bar{X},
\end{aligned}$$

so is PSD. This shows that $\hat{\beta}_{EGMM}$ obtained with $Z = \Psi^T \bar{X}$ is indeed optimal.