LECTURE 12

GMM II

Efficient GMM

The GMM estimator depends on the choice of the weight matrix A_n. The efficient GMM estimator is the estimator with the smallest asymptotic variance among all GMM estimators (defined by different choices of A_n). We show that the efficient GMM estimator corresponds to A_n such that

$$A_n^\top A_n \to_p \Omega^{-1}.$$

Theorem 1. *(a) A lower bound for the asymptotic variance of the class of GMM estimators indexed by A_n is $(Q^\top \Omega^{-1} Q)^{-1}$.*

(b) The lower bound is achieved if $A_n^\top A_n \to_p \Omega^{-1}$.

Proof. To prove part (a), we show that

$$\left(Q^\top \Omega^{-1} Q\right)^{-1} - \left(Q^\top A^\top A Q\right)^{-1} Q^\top A^\top A \Omega A^\top A Q \left(Q^\top A^\top A Q\right)^{-1}$$

is negative semidefinite for any A that has rank l. Equivalently, we can show that

$$Q^\top \Omega^{-1} Q - Q^\top A^\top A Q \left(Q^\top A^\top A \Omega A^\top A Q\right)^{-1} Q^\top A^\top A Q \tag{1}$$

is positive semidefinite.

Since Ω is positive definite, we can write

$$\Omega^{-1} = C^\top C,$$

where C is invertible as well. Rewrite (1) as

$$\begin{aligned} &Q^\top C^\top C Q - Q^\top A^\top A Q \left(Q^\top A^\top A C^{-1} \left(C^\top\right)^{-1} A^\top A Q\right)^{-1} Q^\top A^\top A Q \\ &= Q^\top C^\top \left(I - \left(C^\top\right)^{-1} A^\top A Q \left(Q^\top A^\top A C^{-1} \left(C^\top\right)^{-1} A^\top A Q\right)^{-1} Q^\top A^\top A C^{-1}\right) C Q. \end{aligned} \tag{2}$$

Define

$$H = \left(C^\top\right)^{-1} A^\top A Q,$$

and note that, using this definition, (2) becomes

$$Q^\top C^\top \left(I - H \left(H^\top H\right)^{-1} H^\top\right) C Q.$$

The above matrix is positive semidefinite if $I - H\left(H^\top H\right)^{-1} H^\top$ is positive semidefinite. Next,

$$\begin{aligned} &\left(I - H \left(H^\top H\right)^{-1} H^\top\right)\left(I - H \left(H^\top H\right)^{-1} H^\top\right) \\ &= I - 2H \left(H^\top H\right)^{-1} H^\top + H \left(H^\top H\right)^{-1} H^\top H \left(H^\top H\right)^{-1} H^\top \\ &= I - H \left(H^\top H\right)^{-1} H^\top. \end{aligned}$$

Therefore, $I - H\left(H^\top H\right)^{-1} H^\top$ is symmetric and idempotent, and consequently positive semidefinite. This completes the proof of part (a).

For part (b), if $A_n^\top A_n \to_p A^\top A = \Omega^{-1}$, then the asymptotic variance becomes

$$\begin{aligned}&\left(Q^\top\Omega^{-1}Q\right)^{-1} Q^\top\Omega^{-1}\Omega\Omega^{-1}Q\left(Q^\top\Omega^{-1}Q\right)^{-1}\\&=\left(Q^\top\Omega^{-1}Q\right)^{-1}.\end{aligned}$$

□

A natural choice for such $A_n^\top A_n$ is $\widehat{\Omega}_n^{-1}$. This suggests the following *two-step* procedure:

1. Set $A_n^\top A_n = I_l$. Obtain the corresponding (inefficient) estimator of β, denoted $\widetilde{\beta}_n$. Using the inefficient but consistent estimator of β, obtain $\widehat{\Omega}_n$. For example, in the linear case,

$$\begin{aligned}\widehat{\Omega}_n &= n^{-1}\sum_{i=1}^n \widehat{U}_i^2 Z_i Z_i^\top, \text{ where}\\ \widehat{U}_i &= Y_i - X_i^\top\widetilde{\beta}_n,\end{aligned}$$

   and, in the general case,

$$\widehat{\Omega}_n = n^{-1}\sum_{i=1}^n g\left(W_i,\widetilde{\beta}_n\right) g\left(W_i,\widetilde{\beta}_n\right)^\top.$$

2. Obtain the efficient GMM estimates of β by minimizing

$$\left(n^{-1}\sum_{i=1}^n g(W_i,b)\right)^\top \widehat{\Omega}_n^{-1}\left(n^{-1}\sum_{i=1}^n g(W_i,b)\right),$$

   where $\widehat{\Omega}_n$ is from step 1.

An alternative to $\widehat{\Omega}_n$ in step 1 is

$$n^{-1}\sum_{i=1}^n\left(g\left(W_i,\widetilde{\beta}_n\right) - n^{-1}\sum_{j=1}^n g\left(W_j,\widetilde{\beta}_n\right)\right)\left(g\left(W_i,\widetilde{\beta}_n\right) - n^{-1}\sum_{j=1}^n g\left(W_j,\widetilde{\beta}_n\right)\right)^\top,$$

the centered version of $\widehat{\Omega}_n$. The two versions are asymptotically equivalent because $\mathrm{E}[g(W_i,\beta)] = 0$. However, the centered version often performs better in finite samples.

In the linear case, a better choice for the first-stage weight matrix is

$$\begin{aligned}A_n^\top A_n &= \left(\sum_{i=1}^n Z_i Z_i^\top\right)^{-1}\\ &= \left(Z^\top Z\right)^{-1}.\end{aligned} \tag{3}$$

The reason for this becomes clear in the next section.

The variance-covariance matrix of the efficient GMM estimator can be estimated consistently by

$$\left(\widehat{Q}_n^\top\widehat{\Omega}_n^{-1}\widehat{Q}_n\right)^{-1},$$

where $\widehat{Q}_n$ was defined in Lecture 11. One can use $\widehat{\Omega}_n$ from the first stage or recompute it using the efficient GMM estimator to construct $\widehat{U}_i$ in the linear case or $g(W_i,\widehat{\beta}_n^{GMM})$ in the general case.

Two-stage Least Squares (2SLS)

Consider the linear IV regression model, and assume that

$$\mathrm{E}\left[U_i^2 \mid Z_i\right] = \sigma^2. \tag{4}$$

In this case,

$$\begin{aligned}\Omega &= \mathrm{E}\left[U_i^2 Z_i Z_i^\top\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[U_i^2 \mid Z_i\right] Z_i Z_i^\top\right] \\ &= \sigma^2\, \mathrm{E}\left[Z_i Z_i^\top\right].\end{aligned}$$

A natural estimator of $\mathrm{E}\left[Z_i Z_i^\top\right]$ is

$$n^{-1} \sum_{i=1}^{n} Z_i Z_i^\top,$$

which gives the optimal weight matrix as in (3). Note that, in this case, the efficient GMM estimator can be obtained without the first step, since the weight matrix in (3) does not depend on $\widehat{U}_i$. The efficient GMM estimator is given by

$$\begin{aligned}\widehat{\beta}_n^{2SLS} &= \left(\sum_{i=1}^{n} X_i Z_i^\top \left(\sum_{i=1}^{n} Z_i Z_i^\top\right)^{-1} \sum_{i=1}^{n} Z_i X_i^\top\right)^{-1} \sum_{i=1}^{n} X_i Z_i^\top \left(\sum_{i=1}^{n} Z_i Z_i^\top\right)^{-1} \sum_{i=1}^{n} Z_i Y_i \\ &= \left(X^\top Z \left(Z^\top Z\right)^{-1} Z^\top X\right)^{-1} X^\top Z \left(Z^\top Z\right)^{-1} Z^\top Y.\end{aligned}$$

Its asymptotic distribution is

$$n^{1/2}\left(\widehat{\beta}_n^{2SLS} - \beta\right) \to_d N\left(0, \sigma^2 \left(\mathrm{E}[X_i Z_i^\top]\left(\mathrm{E}[Z_i Z_i^\top]\right)^{-1} \mathrm{E}[Z_i X_i^\top]\right)^{-1}\right).$$

This estimator is also called the two-stage least-squares estimator for the following reason. Define

$$\begin{aligned}\widetilde{X} &= Z\left(Z^\top Z\right)^{-1} Z^\top X \\ &= P_Z X,\end{aligned}$$

the orthogonal projection of X onto the column space of Z. Since P_Z is idempotent, we can write

$$\widehat{\beta}_n^{2SLS} = \left(\widetilde{X}^\top \widetilde{X}\right)^{-1} \widetilde{X}^\top Y.$$

The estimator $\widehat{\beta}_n^{2SLS}$ can be obtained via the following two-step procedure. In the first step, regress X on the instruments to obtain fitted values $\widetilde{X}$. This replaces the endogenous regressors with fitted values that lie in the column space of Z and are uncorrelated with U_i. In the second step, regress Y on $\widetilde{X}$.

The 2SLS estimator is not efficient when the conditional homoskedasticity assumption (4) fails. In this case, the efficient GMM estimator is

$$\widehat{\beta}_n^{GMM} = \left(\sum_{i=1}^{n} X_i Z_i^\top \left(\sum_{i=1}^{n} \widehat{U}_i^2 Z_i Z_i^\top\right)^{-1} \sum_{i=1}^{n} Z_i X_i^\top\right)^{-1} \sum_{i=1}^{n} X_i Z_i^\top \left(\sum_{i=1}^{n} \widehat{U}_i^2 Z_i Z_i^\top\right)^{-1} \sum_{i=1}^{n} Z_i Y_i.$$

Exactly identified case

When the number of instruments is equal to the number of regressors ($l = k$), and the $k \times k$ matrix $Z^\top X$ is of full rank, the 2SLS estimator reduces to the IV estimator discussed in Lecture 10:

$$\begin{aligned}
\widehat{\beta}_n^{2SLS} &= \left(X^\top Z \left(Z^\top Z\right)^{-1} Z^\top X\right)^{-1} X^\top Z \left(Z^\top Z\right)^{-1} Z^\top Y \\
&= \left(Z^\top X\right)^{-1} \left(Z^\top Z\right) \left(X^\top Z\right)^{-1} X^\top Z \left(Z^\top Z\right)^{-1} Z^\top Y \\
&= \left(Z^\top X\right)^{-1} Z^\top Y \\
&= \widehat{\beta}_n^{IV}.
\end{aligned}$$

The IV estimator is a linear example of the exactly identified case. In this case, the weight matrix A_n plays no role. If the model is exactly identified, then we have k equations in k unknowns. Therefore, it is possible to solve $n^{-1} \sum_{i=1}^n g(W_i, b) = 0$ exactly. As a result, the solution to the GMM minimization problem

$$\min_{b \in B} \left\| A_n n^{-1} \sum_{i=1}^n g(W_i, b) \right\|^2$$

does not depend on A_n.

Since, in the exactly identified case, Q is $k \times k$ and invertible, the asymptotic variance-covariance matrix takes the following form:

$$\begin{aligned}
&\left(Q^\top A^\top A Q\right)^{-1} Q^\top A^\top A \Omega A^\top A Q \left(Q^\top A^\top A Q\right)^{-1} \\
&= Q^{-1} \left(A^\top A\right)^{-1} \left(Q^\top\right)^{-1} Q^\top A^\top A \Omega A^\top A Q Q^{-1} \left(A^\top A\right)^{-1} \left(Q^\top\right)^{-1} \\
&= Q^{-1} \Omega \left(Q^{-1}\right)^\top \\
&= \left(Q^\top \Omega^{-1} Q\right)^{-1},
\end{aligned}$$

independent of A and therefore efficient.

Confidence intervals and hypothesis testing in the GMM framework

In this section, we discuss the construction of confidence intervals and hypothesis testing. Let $\widehat{\beta}_n^{GMM}$ be the efficient GMM estimator with the asymptotic variance-covariance matrix $V = \left(Q^\top \Omega^{-1} Q\right)^{-1}$. Let $\widehat{V}_n$ denote a consistent estimator of V.

Since $\widehat{\beta}_n^{GMM}$ is approximately normal in large samples, a confidence interval with the nominal coverage probability $1 - \alpha$ for element j of β is given by

$$\left[\widehat{\beta}_{n,j}^{GMM} - z_{1-\alpha/2} \sqrt{\left[\widehat{V}_n\right]_{jj} / n}, \widehat{\beta}_{n,j}^{GMM} + z_{1-\alpha/2} \sqrt{\left[\widehat{V}_n\right]_{jj} / n}\right],$$

for $j = 1, \ldots, k$.

For example, in the linear and homoskedastic case, the asymptotic variance of $\widehat{\beta}_n^{2SLS}$ is

$$V = \sigma^2 \left(\mathrm{E}[X_i Z_i^\top] \left(\mathrm{E}[Z_i Z_i^\top]\right)^{-1} \mathrm{E}[Z_i X_i^\top]\right)^{-1},$$

and its consistent estimator is

$$\begin{aligned}
\widehat{V}_n &= \widehat{\sigma}_n^2 \left(n^{-1} \sum_{i=1}^n X_i Z_i^\top \left(n^{-1} \sum_{i=1}^n Z_i Z_i^\top\right)^{-1} n^{-1} \sum_{i=1}^n Z_i X_i^\top\right)^{-1} \\
&= n \widehat{\sigma}_n^2 \left(X^\top Z \left(Z^\top Z\right)^{-1} Z^\top X\right)^{-1},
\end{aligned}$$

where $\widehat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \left(Y_i - X_i^\top \widehat{\beta}_n^{2SLS} \right)^2$. Therefore, the asymptotic $1-\alpha$ confidence interval for β_j is given by

$$\widehat{\beta}_{n,j}^{2SLS} \pm z_{1-\alpha/2} \sqrt{\widehat{\sigma}_n^2 \left[\left(X^\top Z \left(Z^\top Z \right)^{-1} Z^\top X \right)^{-1} \right]_{jj}}.$$

One can construct a test of the null hypothesis $H_0 : \beta_j = \beta_{0,j}$ against $H_1 : \beta_j \neq \beta_{0,j}$ by using the following test statistic:

$$T_{n,j} = \frac{\widehat{\beta}_{n,j}^{GMM} - \beta_{0,j}}{\sqrt{\left[\widehat{V}_n \right]_{jj} / n}}.$$

Since, under the null hypothesis, $T_{n,j} \to_d N(0,1)$, the asymptotic α-size test is given by

$$\text{Reject } H_0 \text{ if } |T_{n,j}| > z_{1-\alpha/2}.$$

A Wald statistic can be used to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$:

$$W_n = n \left(\widehat{\beta}_n^{GMM} - \beta_0 \right)^\top \widehat{V}_n^{-1} \left(\widehat{\beta}_n^{GMM} - \beta_0 \right).$$

More generally, suppose that the null and alternative are given by $H_0 : h(\beta) = 0$ and $H_1 : h(\beta) \neq 0$, where $h : \mathbb{R}^k \to \mathbb{R}^q$. By the delta method, under the null hypothesis,

$$n^{1/2} h(\widehat{\beta}_n^{GMM}) \to_d N \left(0, \frac{\partial h(\beta)}{\partial \beta^\top} V \left(\frac{\partial h(\beta)}{\partial \beta^\top} \right)^\top \right).$$

Therefore, the Wald statistic is given by

$$W_n = n h(\widehat{\beta}_n^{GMM})^\top \left(\frac{\partial h(\widehat{\beta}_n^{GMM})}{\partial \beta^\top} \widehat{V}_n \left(\frac{\partial h(\widehat{\beta}_n^{GMM})}{\partial \beta^\top} \right)^\top \right)^{-1} h(\widehat{\beta}_n^{GMM}).$$

The asymptotic α-size test is given by

$$\text{Reject } H_0 \text{ if } W_n > \chi^2_{q,1-\alpha}.$$

Testing overidentified restrictions

In this section, we discuss a *specification test* for whether the moment condition $\mathrm{E}[g(W_i, \beta)] = 0$ holds. Unlike the tests discussed before, this is not a test of whether β takes a specific value but rather whether the model, as defined by the moment conditions, is correctly specified. The null hypothesis is that there exists some β such that $\mathrm{E}[g(W_i, \beta)] = 0$. The alternative hypothesis is that $\mathrm{E}[g(W_i, \beta)] \neq 0$ for all $\beta \in \mathbb{R}^k$. Note that, when the model is exactly identified, the system of k equations in k unknowns $\mathrm{E}[g(W_i, b)] = 0$ can be solved exactly. Thus, we can test the validity of the moment restrictions only if the model is overidentified.

When the model is overidentified, in general, it is impossible to choose b such that $n^{-1} \sum_{i=1}^n g(W_i, b)$ is exactly zero. However, if the moment condition $\mathrm{E}[g(W_i, \beta)] = 0$ holds, we should expect that $n^{-1} \sum_{i=1}^n g(W_i, \beta)$ is close to zero, and further,

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n g(W_i, \beta) &\to_d N \left(0, \mathrm{E}[g(W_i, \beta) g(W_i, \beta)^\top] \right) \\ &= N(0, \Omega). \end{aligned}$$

If we use the efficient matrix A_n, then

$$A_n^\top A_n \to_p \Omega^{-1}. \tag{5}$$

In this case, the weighted distance

$$\left(n^{-1/2}\sum_{i=1}^{n} g(W_i,\beta)\right)^\top A_n^\top A_n \left(n^{-1/2}\sum_{i=1}^{n} g(W_i,\beta)\right)$$

asymptotically has the χ_l^2 distribution (the degrees of freedom are determined by the l moment restrictions). When β is replaced by its *efficient* GMM estimator $\widehat{\beta}_n^{GMM}$, the degrees of freedom change from l to $l-k$. We have the following result. Under the null hypothesis $H_0 : \mathrm{E}[g(W_i,\beta)] = 0$ for some $\beta \in \mathbb{R}^k$, and provided that A_n satisfies (5) and $\widehat{\beta}_n^{GMM}$ is efficient,

$$\left(n^{-1/2}\sum_{i=1}^{n} g\left(W_i,\widehat{\beta}_n^{GMM}\right)\right)^\top A_n^\top A_n \left(n^{-1/2}\sum_{i=1}^{n} g\left(W_i,\widehat{\beta}_n^{GMM}\right)\right) \to_d \chi_{l-k}^2.$$

The reason for the change in degrees of freedom is that we have to estimate k parameters β before constructing the test statistic. Another explanation is that we need k restrictions to estimate β. Thus, we can test only the additional (overidentified) $l-k$ restrictions.

Consider the linear and homoskedastic case. The efficient GMM estimator is the 2SLS estimator, and the efficient weight matrix is given by $\left(\sum_{i=1}^n Z_i Z_i^\top\right)^{-1}$. One should reject the null of a correctly specified model if

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n} \widehat{U}_i Z_i^\top \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i^\top\right)^{-1} n^{-1/2}\sum_{i=1}^{n} Z_i \widehat{U}_i/\widehat{\sigma}_n^2 \\
&= \left(\sum_{i=1}^{n}\left(Y_i - X_i^\top \widehat{\beta}_n^{GMM}\right) Z_i\right)^\top \left(\sum_{i=1}^{n} Z_i Z_i^\top\right)^{-1} \left(\sum_{i=1}^{n}\left(Y_i - X_i^\top \widehat{\beta}_n^{GMM}\right) Z_i\right)/\widehat{\sigma}_n^2 \\
&> \chi_{l-k,1-\alpha}^2,
\end{aligned}$$

where $\widehat{\sigma}_n^2$ is any consistent estimator of $\sigma^2 = \mathrm{E}[U_i^2]$, such as $n^{-1}\sum_{i=1}^n \left(Y_i - X_i^\top \widehat{\beta}_n^{GMM}\right)^2$. Here, we test *jointly* the exogeneity of the instruments and other assumptions, such as linearity of the model.