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 one that has the smallest asymptotic variance among all GMM estimators (defined by different choices of A_n). Next, we will show that the efficient GMM corresponds to A_n such that

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

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

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

Proof. In order to prove part (a), we need to show that

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

is negative semi-definite for any A that has rank l. Equivalently, we can show that

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

is positive semi-definite.

Since the inverse of Ω exists (Ω is positive definite), we can write

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

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

$$\begin{aligned} & Q'C'CQ - Q'A'AQ\left(Q'A'AC^{-1}\left(C'\right)^{-1}A'AQ\right)^{-1}Q'A'AQ \\ = \;& Q'C'\left(I - \left(C'\right)^{-1}A'AQ\left(Q'A'AC^{-1}\left(C'\right)^{-1}A'AQ\right)^{-1}Q'A'AC^{-1}\right)CQ. \end{aligned} \tag{2}$$

Define

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

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

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

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

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

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

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

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

□

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

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

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

   and, in the general case,

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

2. Obtain the efficient GMM estimates of β by minimizing

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

   where $\widehat{\Omega}_n$ comes from the first step.

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

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

the centered version of $\widehat{\Omega}_n$. The two versions are asymptotically equivalent, since $E\partial g(W_i, \beta)/\partial b' = 0$. However, the centered version often performs better.

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

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

The reason for this become clear in the next section.

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

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

where $\widehat{Q}_n$ was defined in Lecture 11. One can use $\widehat{\Omega}_n$ from the first stage, or compute it again, using the efficient GMM estimator to compute $\widehat{U}_i$'s in the linear case or $\partial g/\partial b'$ in the general case.

Two-stage Least Squares (2SLS)

Consider the linear IV regression model, and assume that

$$E\left(U_i^2 | Z_i\right) = \sigma^2. \qquad (4)$$

In this case,

$$\begin{aligned} \Omega &= E\left(U_i^2 Z_i Z_i'\right) \\ &= E\left(E\left(U_i^2 | Z_i\right) Z_i Z_i'\right) \\ &= \sigma^2 E\left(Z_i Z_i'\right). \end{aligned}$$

A natural estimator of $E\left(Z_i Z_i'\right)$ is
$$n^{-1}\sum_{i=1}^{n} Z_i Z_i',$$
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 $\widehat{U}_i$'s. The efficient GMM is given by
$$\begin{aligned}\widehat{\beta}_n^{2SLS} &= \left(\sum_{i=1}^{n} X_i Z_i' \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i Z_i' \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \sum_{i=1}^{n} Z_i Y_i \\ &= \left(X'Z\left(Z'Z\right)^{-1} Z'X\right)^{-1} X'Z\left(Z'Z\right)^{-1} Z'Y.\end{aligned}$$
We have that
$$n^{1/2}\left(\widehat{\beta}_n^{2SLS}-\beta\right) \to_d N\left(0, \sigma^2\left(EX_iZ_i'\left(EZ_iZ_i'\right)^{-1} EZ_iX_i'\right)^{-1}\right).$$

The above estimator is also called the two stage LS estimator for the following reason. Define
$$\begin{aligned}\widetilde{X} &= Z\left(Z'Z\right)^{-1} Z'X \\ &= P_Z X,\end{aligned}$$
the orthogonal projection of the matrix of regressors X onto the space spanned by the instruments Z. Since P_Z is idempotent, we can write
$$\widehat{\beta}_n^{2SLS} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'Y.$$
Thus, $\widehat{\beta}_n$ can be obtained using the two-step procedure. First, regress X against instruments, and obtain the fitted values $\widetilde{X}$. The first step removes from X_i the correlation with the error U_i. In the second step, one should run the regression of Y against the fitted values $\widetilde{X}$.

The 2SLS estimator is not efficient when the conditional homoskedasticity assumption (4) fails. In this case, the efficient GMM estimator is
$$\widehat{\beta}_n^{GMM} = \left(\sum_{i=1}^{n} X_i Z_i' \left(\sum_{i=1}^{n} \widehat{U}_i^2 Z_i Z_i'\right)^{-1} \sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i Z_i' \left(\sum_{i=1}^{n} \widehat{U}_i^2 Z_i Z_i'\right)^{-1} \sum_{i=1}^{n} Z_i Y_i$$

Exactly identified case

When the number of instruments is equal to the number of regressors ($l = k$), and the $k \times k$ matrix $Z'X$ is of full rank, the 2SLS estimator reduces to the IV estimator discussed in Lecture 10:
$$\begin{aligned}\widehat{\beta}_n^{2SLS} &= \left(X'Z\left(Z'Z\right)^{-1} Z'X\right)^{-1} X'Z\left(Z'Z\right)^{-1} Z'Y \\ &= \left(Z'X\right)^{-1}\left(Z'Z\right)\left(X'Z\right)^{-1} X'Z\left(Z'Z\right)^{-1} Z'Y \\ &= \left(Z'X\right)^{-1} Z'Y \\ &= \widehat{\beta}_n^{IV}.\end{aligned}$$

The IV estimator is an example (linear) of the exactly identified case. In this case, the weight matrix A_n plays no role. If the model is exactly identified, the we have k equations in k unknowns. Therefore, it is possible to solve $n^{-1}\sum_{i=1}^{n} g\left(W_i, b\right) = 0$ exactly. As a result, the solution to the GMM minimization problem
$$\min_{b\in B}\left\|A_n n^{-1}\sum_{i=1}^{n} g\left(W_i, b\right)\right\|^2$$

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

$$\begin{aligned} & \left(Q'A'AQ\right)^{-1} Q'A'A\Omega A'AQ \left(Q'A'AQ\right)^{-1} \\ = \; & Q^{-1}\left(A'A\right)^{-1}\left(Q'\right)^{-1} Q'A'A\Omega A'AQQ^{-1}\left(A'A\right)^{-1}\left(Q'\right)^{-1} \\ = \; & Q^{-1}\Omega\left(Q^{-1}\right)' \\ = \; & \left(Q'\Omega^{-1}Q\right)^{-1} \end{aligned}$$

independent of A and, naturally, efficient.

Confidence intervals and hypothesis testing in the GMM framework

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

$$V = \sigma^2\left(EX_iZ_i'\left(EZ_iZ_i'\right)^{-1}EZ_iX_i'\right)^{-1},$$

and its consistent estimator is

$$\begin{aligned} \widehat{V}_n &= \widehat{\sigma}_n^2\left(n^{-1}\sum_{i=1}^n X_iZ_i'\left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right)^{-1} n^{-1}\sum_{i=1}^n Z_iX_i'\right)^{-1} \\ &= n\widehat{\sigma}_n^2\left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}, \end{aligned}$$

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

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

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

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

One can use a Wald statistic in order to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$:

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

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

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

Therefore, the Wald statistic is given by

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

The asymptotic α-size test is given by

$$\text{Reject } H_0 \text{ if } W_n > \chi_q^2.$$

Testing overidentified restrictions

In this section, we discuss a *specification test* that allows one to test whether the moment condition $Eg(W_i, \beta) = 0$. Contrary to the tests discussed before, this is not a test of whether β takes on some 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 $Eg(W_i, \beta) = 0$. The alternative hypothesis is that $Eg(W_i, \beta) \neq 0$ for all $\beta \in R^k$. Note that, when the model is exactly identified, the system of k equations in k unknowns $Eg(W_i, b) = 0$ can be solved exactly. Thus, we can test validity of 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 $Eg(W_i, \beta) = 0$ holds, we should expect that $n^{-1}\sum_{i=1}^n g(W_i, \beta)$ is close to zero, and further,

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

If we use the efficient matrix A_n, then

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

In this case, the weighted distance

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

asymptotically has the χ_l^2 distribution (the degrees of freedom are determined by the l moment restrictions). It turns out that, 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 $H_0 : Eg(W_i, \beta) = 0$ for some $\beta \in R^k$, and provided that A_n satisfies (5) and $\widehat{\beta}_n^{GMM}$ is efficient,

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

The reason for change in degrees of freedom is that we have to estimate k parameters β before construction the test statistic. Another explanation is that we need k restrictions to estimate β. Thus, we can test only 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 $\left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}$. One should reject the null of correctly specified model if

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

where $\widehat{\sigma}_n^2$ is any consistent estimator of $\sigma^2 = EU_i^2$, such as $n^{-1} \sum_{i=1}^{n} \left(Y_i - X_i' \widehat{\beta}_n^{GMM} \right)^2$. Note that here we test *jointly* exogeneity of the instruments and other assumptions such as linearity of the model.