GMM I

Definition

Suppose that an econometrician observes the data $\{W_i : i = 1, \ldots, n\}$ where W_i is a random p-vector. Let g be a l dimensional function depending on W_i and the k-vector of parameters b:

$$g(W_i, b) = \begin{pmatrix} g_1(W_i, b) \\ \vdots \\ g_l(W_i, b) \end{pmatrix},$$

and $g_j : R^p \times R^k \to R$ for $j = 1, \ldots, l$. The model is defined by the following *moment condition.*

$$Eg(W_i, \beta) = 0 \text{ for some } \beta \in R^k. \tag{1}$$

Examples:

- **Linear regression.** Let $W_i = (Y_i, X_i')'$ and $Y_i = X_i'\beta + U_i$, where $\beta \in R^k$, and $E(X_i U_i) = 0$. In this case, $g(W_i, b) = X_i(Y_i - X_i'b)$, $l = k$, and the moment condition is $E(X_i(Y_i - X_i'\beta)) = 0$.
- **IV regression.** Let $W_i = (Y_i, X_i', Z_i')'$, $Y_i = X_i'\beta + U_i$, where $\beta \in R^k$, and $E(Z_i U_i) = 0$, where Z_i is a l-vector. In this case, $g(W_i, b) = Z_i(Y_i - X_i'b)$ with the moment condition $E(Z_i(Y_i - X_i'\beta)) = 0$.
- **Lucas' Model.** Suppose that in period t investors receive utility from the consumption C_t be consumption in period t. Let $R_{j,t}$ be the rate of return on the risky asset j. Suppose that there are m assets. Assume that the utility function is of the form $\sum_{t=1}^{\infty} \delta^t C_t^{1-\alpha}/(1-\alpha)$. In the equilibrium, the returns on risky assets are determined by the following Euler equations:

  $$E\left(\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\alpha}(1 + R_{j,t+1})\right) = 1 \text{ for } j = 1, \ldots, m.$$

  In this case we have $W_t = (C_t, R_{1,t} \ldots, R_{m,t})$, $b = (a, d)$, $g_j(W_t, b) = d\left(\frac{C_{t+1}}{C_t}\right)^{-a}(1 + R_{j,t+1}) - 1$ for $j = 1, \ldots, m$, and the moment conditions given by the above equations. Note that in this case g is nonlinear in the parameters.

We say that the model is *identified* if $Eg(W_i, \beta) = 0$ and $Eg\left(W_i, \widetilde{\beta}\right) = 0$ imply that $\beta = \widetilde{\beta}$, i.e. the solution of (1) is unique. The moment condition gives us l restrictions for k parameters. A necessary condition for the model to be identified is that $l \geq k$, i.e. we must have *at least* k restrictions. The necessary condition is called the *order condition.* We say that the model is not identified or *underidentified* if the order condition fails.

When $k = l$, applying the MM principle, we can estimate β by the value of b that solves the sample analogue of (1):

$$n^{-1}\sum_{i=1}^{n} g\left(W_i, \widehat{\beta}_n^{MM}\right) = 0.$$

However, when $l > k$, in general, there is no $b \in R^k$ that solves all l equations exactly. In this case, we can choose the value of b that makes the sample moments as close to zero as possible. Let A_n be a (possibly random) $l \times l$ weight matrix such that $A_n \to_p A$, where A is non-random and has full rank (l). The *Generalized Method of Moments (GMM) estimator* of β is defined to be the value of b that minimizes the weighted distance of $n^{-1}\sum_{i=1}^{n} g(W_i, b)$ from zero:

$$\begin{aligned} \widehat{\beta}_n^{GMM} &= \arg\min_{b\in B} \left\| A_n n^{-1} \sum_{i=1}^{n} g(W_i, b) \right\|^2 \\ &= \arg\min_{b\in B} \left(n^{-1} \sum_{i=1}^{n} g(W_i, b) \right)' A_n' A_n \left(n^{-1} \sum_{i=1}^{n} g(W_i, b) \right). \end{aligned} \tag{2}$$

The set $B \subset R^k$ is usually assumed to be compact. Note that $A'A$ is positive definite.

Linear case

In this section, we discuss in details the IV regression example. Note that in this case, the function g is linear in parameters. Similarly to Lecture 10, we assume that some or all of the k regressors in X_i are endogenous:

$$E(X_i U_i) \neq 0,$$

and that the l instruments Z_i are weakly exogenous:

$$E(Z_i U_i) = 0.$$

The model is identified, if the following *rank condition* is satisfied:

$$rank\left(E\left(Z_i X_i'\right)\right) = k.$$

If the rank condition is satisfied and $l = k$ we say that the model is *exactly* or *just identified.* We say that the model is *overidentified* if the rank condition is satisfied and $l > k$ (there are more instruments than the parameters that want to estimate). Contrary to the discussion in Lecture 10, we allow here the model to be overidentified.

In the linear IV regression case, $\widehat{\beta}_n^{GMM}$ is the minimizer of

$$\left(n^{-1} \sum_{i=1}^{n} Z_i (Y_i - X_i' b) \right)' A_n' A_n \left(n^{-1} \sum_{i=1}^{n} Z_i (Y_i - X_i' b) \right),$$

and given by the following expression:

$$\widehat{\beta}_n^{GMM} = \left(\sum_{i=1}^{n} X_i Z_i' (A_n' A_n) \sum_{i=1}^{n} Z_i X_i' \right)^{-1} \sum_{i=1}^{n} X_i Z_i' (A_n' A_n) \sum_{i=1}^{n} Z_i Y_i.$$

We will show next that the GMM estimator is consistent. We need the following assumptions.

- $\{(Y_i, X_i, Z_i) : i = 1, \ldots, n\}$ are iid.
- $Y_i = X_i'\beta + U_i$, where $\beta \in R^k$.
- $E(Z_i U_i) = 0$.
- $E(Z_i X_i')$ has rank k.
- $A_n \to_p A$, where A has rank $l \geq k$.
- $EX_{i,j}^2 < \infty$ for all $j = 1, \ldots, k$.
- $EZ_{i,j}^2 < \infty$ for all $j = 1, \ldots, l$.

Write

$$\widehat{\beta}_n^{GMM} = \beta + \left(n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1} \sum_{i=1}^{n} Z_i X_i' \right)^{-1} n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1} \sum_{i=1}^{n} Z_i U_i.$$

The last two of the above assumptions imply that

$$E \left| X_{i,r} Z_{i,s} \right| < \infty \text{ for all } r = 1, \ldots, k \text{ and } s = 1, \ldots, l.$$

By the WLLN,

$$n^{-1} \sum_{i=1}^{n} X_i Z_i' \to_p E X_i Z_i'.$$

Since $A_n \to_p A$, we also have that

$$n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1} \sum_{i=1}^{n} Z_i X_i' \to_p E X_i Z_i' \left(A' A \right) E Z_i X_i'.$$

Further, since $E \left(Z_i X_i' \right)$ has rank k, A has rank $l \geq k$, it follows that the $k \times k$ matrix $E X_i Z_i' \left(A' A \right) E Z_i X_i'$ has full rank k and, therefore, invertible. Consequently, by the Slutsky's Theorem,

$$\left(n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1} \sum_{i=1}^{n} Z_i X_i' \right)^{-1} \to_p \left(E X_i Z_i' \left(A' A \right) E Z_i X_i' \right)^{-1}.$$

Next, by the WLLN,

$$n^{-1} \sum_{i=1}^{n} Z_i U_i \to_p 0,$$

and thus $\widehat{\beta}_n^{GMM} \to_p \beta$.

In order to show asymptotic normality, we will need the following two assumptions in addition to the above.

- $E Z_{i,j}^4 < \infty$ for all $j = 1, \ldots, l$.
- $E U_i^4 < \infty$.
- $E U_i^2 Z_i Z_i'$ is positive definite.

Write

$$n^{1/2} \left(\widehat{\beta}_n^{GMM} - \beta \right) = \left(n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1} \sum_{i=1}^{n} Z_i X_i' \right)^{-1} n^{-1} \sum_{i=1}^{n} X_i Z_i' \left(A_n' A_n \right) n^{-1/2} \sum_{i=1}^{n} Z_i U_i.$$

The last two assumptions imply that the variance of $Z_i U_i$, $E U_i^2 Z_i Z_i'$ is finite, and, by the CLT, we have that

$$n^{-1/2} \sum_{i=1}^{n} Z_i U_i \to_d N \left(0, E U_i^2 Z_i Z_i' \right).$$

Let's define

$$\begin{aligned} Q &= E Z_i X_i', \\ \Omega &= E U_i^2 Z_i Z_i'. \end{aligned}$$

Combining the above results, we have

$$n^{1/2}\left(\widehat{\beta}_n^{GMM}-\beta\right)\to_d N\left(0,V\right),$$

where V takes the sandwich form:

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

The variance-covariance matrix V can be estimated by replacing A, Q and Ω with their consistent estimators A_n, $\widehat{Q}_n$ and $\widehat{\Omega}_n$ respectively, where

$$\begin{aligned}\widehat{Q}_n &= n^{-1}\sum_{i=1}^n Z_iX_i',\\ \widehat{\Omega}_n &= n^{-1}\sum_{i=1}^n \widehat{U}_i^2 Z_iZ_i',\end{aligned}$$

where $\widehat{U}_i=Y_i-X_i'\widehat{\beta}_n^{GMM}$.

General case

In the general case, the GMM estimator minimizes the nonlinear function in (2). Usually, we do not have a closed-form expression for $\widehat{\beta}_n^{GMM}$, and the minimization must be done using numerical procedures. Nevertheless, under general regularity conditions, it is possible to show that $\widehat{\beta}_n^{GMM}$ is consistent and asymptotically normal. We will only provide heuristic proofs of consistency and asymptotic normality.

Since the criterion function in (2) involves averages, we should expect that

$$\left\|A_n n^{-1}\sum_{i=1}^n g\left(W_i,b\right)\right\|^2\to_p\left\|AEg\left(W_i,b\right)\right\|^2. \tag{3}$$

Assuming that the model is uniquely identified, $Eg\left(W_i,b\right)=0$ only if and only if $b=\beta$. Since $\left\|AEg\left(W_i,b\right)\right\|^2>0$ for all $b\neq\beta$, the true value β is the unique minimizer of $\left\|AEg\left(W_i,b\right)\right\|^2$. Intuitively, $\widehat{\beta}_n^{GMM}$ is consistent because

$$\begin{aligned}\widehat{\beta}_n^{GMM} &= \arg\min_{b\in B}\left\|A_n n^{-1}\sum_{i=1}^n g\left(W_i,b\right)\right\|^2\\ &\to_p \arg\min_{b\in B}\left\|AEg\left(W_i,b\right)\right\|^2\\ &= \beta.\end{aligned}$$

The formal proof of consistency requires a number of regularity conditions, such as *uniform* in $b\in B$ convergence in (3), compactness of B, β being the interior point of B.

For asymptotic normality, note that $\widehat{\beta}_n^{GMM}$ solves the first-order conditions:

$$\left(n^{-1}\sum_{i=1}^n\frac{\partial g\left(W_i,\widehat{\beta}_n^{GMM}\right)}{\partial b'}\right)'A_n'A_n n^{-1}\sum_{i=1}^n g\left(W_i,\widehat{\beta}_n^{GMM}\right)=0. \tag{4}$$

(In fact, it is sufficient if $\widehat{\beta}_n^{GMM}$ solves the first-order conditions approximately, i.e. on the right-hand side of the above equation, instead of zero, we can have a term that goes to zero in probability at the rate $n^{1/2}$.)

Next, using the expansion of $g\left(W_i, \widehat{\beta}_n^{GMM}\right)$ around $g(W_i, \beta)$ (the element-by-element mean value theorem), we obtain

$$g\left(W_i, \widehat{\beta}_n^{GMM}\right) = g(W_i, \beta) + \frac{\partial g\left(W_i, \widehat{\beta}_n^{*}\right)}{\partial b'}\left(\widehat{\beta}_n^{GMM} - \beta\right), \tag{5}$$

where $\widehat{\beta}_n^{*}$ is between $\widehat{\beta}_n^{GMM}$ and β. Substitution of (5) into (4) gives

$$\begin{aligned}
0 \;=\;& \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'}\right)' A_n' A_n n^{-1}\sum_{i=1}^{n} g(W_i, \beta) \\
&+ \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'}\right)' A_n' A_n \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{*}\right)}{\partial b'}\right)\left(\widehat{\beta}_n^{GMM} - \beta\right),
\end{aligned}$$

We can write

$$\begin{aligned}
& n^{1/2}\left(\widehat{\beta}_n^{GMM} - \beta\right) \\
=\;& -\left(\left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'}\right)' A_n' A_n \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{*}\right)}{\partial b'}\right)\right)^{-1} \\
& \times \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'}\right)' A_n' A_n n^{-1/2}\sum_{i=1}^{n} g(W_i, \beta).
\end{aligned}$$

Since $Eg(W_i, \beta) = 0$, we should expect that, under some regularity conditions,

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

(Note that the asymptotic variance depends on the unknown β). Since $\widehat{\beta}_n^{GMM}$ is consistent, and, as a result $\widehat{\beta}_n^{*} \to_p \beta$ as well, we should expect, that under some proper regularity conditions,

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'} &\to_p E\frac{\partial g(W_i, \beta)}{\partial b'}, \\
n^{-1}\sum_{i=1}^{n}\frac{\partial g\left(W_i, \widehat{\beta}_n^{*}\right)}{\partial b'} &\to_p E\frac{\partial g(W_i, \beta)}{\partial b'},
\end{aligned}$$

and that the matrix

$$\left(E\frac{\partial g(W_i, \beta)}{\partial b'}\right)' A'A\left(E\frac{\partial g(W_i, \beta)}{\partial b'}\right)$$

is invertible. Then,

$$n^{1/2}\left(\widehat{\beta}_n^{GMM} - \beta\right) \to_d N(0, V),$$

where

$$\begin{aligned}
V &= \left(Q'A'AQ\right)^{-1} Q'A'A\Omega A'AQ\left(Q'A'AQ\right)^{-1}, \\
Q &= E\frac{\partial g(W_i, \beta)}{\partial b'}, \\
\Omega &= Eg(W_i, \beta)\, g(W_i, \beta)'.
\end{aligned}$$

The variance-covariance matrix V can be estimated by replacing A, Q and Ω with their consistent estimators A_n and

$$\begin{aligned} \widehat{Q}_n &= n^{-1} \sum_{i=1}^{n} \frac{\partial g\left(W_i, \widehat{\beta}_n^{GMM}\right)}{\partial b'}, \\ \widehat{\Omega}_n &= n^{-1} \sum_{i=1}^{n} g\left(W_i, \widehat{\beta}_n^{GMM}\right) g\left(W_i, \widehat{\beta}_n^{GMM}\right)'. \end{aligned}$$