LECTURE 11

GMM I

Definition

Suppose that an econometrician observes data $\{W_i : i = 1, \ldots, n\}$, where W_i is a random p-vector. Let g be an 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 $g_j : \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}$ for $j = 1, \ldots, l$. The model is defined by the following *moment condition*:

$$\mathrm{E}[g(W_i, \beta)] = 0 \text{ for some } \beta \in \mathbb{R}^k. \tag{1}$$

Examples:

- **Linear regression.** Let $W_i = \left(Y_i, X_i^\top\right)^\top$ and $Y_i = X_i^\top \beta + U_i$, where $\beta \in \mathbb{R}^k$, and $\mathrm{E}[X_i U_i] = 0$. In this case, $g(W_i, b) = X_i \left(Y_i - X_i^\top b\right)$, $l = k$, and the moment condition is $\mathrm{E}\left[X_i \left(Y_i - X_i^\top \beta\right)\right] = 0$.
- **IV regression.** Let $W_i = \left(Y_i, X_i^\top, Z_i^\top\right)^\top$, $Y_i = X_i^\top \beta + U_i$, where $\beta \in \mathbb{R}^k$, and $\mathrm{E}[Z_i U_i] = 0$, where Z_i is an l-vector. In this case, $g(W_i, b) = Z_i \left(Y_i - X_i^\top b\right)$ with the moment condition $\mathrm{E}\left[Z_i \left(Y_i - X_i^\top \beta\right)\right] = 0$.
- **Lucas's model.** (This example uses time subscripts t rather than i, reflecting the time-series nature of the data.) Suppose that in period t, investors derive utility from consumption C_t. Let $R_{j,t}$ be the rate of return on risky asset j. Suppose that there are m assets. Assume that the utility function is $\sum_{t=1}^{\infty} \delta^t C_t^{1-\alpha}/(1-\alpha)$. In equilibrium, the returns on risky assets are determined by the following Euler equations:
$$\mathrm{E}\left[\delta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} (1 + R_{j,t+1})\right] = 1, \quad j = 1, \ldots, m.$$
Here $W_t = (C_t, R_{1,t}, \ldots, R_{m,t})$, $b = (a, d)$, $g_j(W_t, b) = d\left(\frac{C_{t+1}}{C_t}\right)^{-a} (1 + R_{j,t+1}) - 1$ for $j = 1, \ldots, m$, and the moment conditions are given by the above equations. Here, g is nonlinear in the parameters.

The model is *identified* if $\mathrm{E}[g(W_i, \beta)] = 0$ and $\mathrm{E}[g(W_i, \widetilde{\beta})] = 0$ imply $\beta = \widetilde{\beta}$, that is, 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 $l \geq k$, that is, there must be at least k restrictions. This necessary condition is called the *order condition*. The model is *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):

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

When $l > k$, no $b \in \mathbb{R}^k$ generally 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:

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

The set $B \subset \mathbb{R}^k$ is assumed to be compact. The matrix $A^\top A$ is positive definite.

Linear case

In this section, we discuss the IV regression example in detail. Here g is linear in the parameters. As in Lecture 10, we assume that some or all of the k regressors in X_i are endogenous:

$$
\mathrm{E}[X_i U_i] \neq 0,
$$

and that the l instruments Z_i are weakly exogenous:

$$
\mathrm{E}[Z_i U_i] = 0.
$$

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

$$
\mathrm{rank}\left(\mathrm{E}\left[Z_i X_i^\top\right]\right) = k.
$$

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 parameters). Unlike in Lecture 10, the model may be overidentified here.

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

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

and is given by the following expression:

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

We 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.
- $Y_i = X_i^\top \beta + U_i$, where $\beta \in \mathbb{R}^k$.
- $\mathrm{E}\left[Z_i U_i\right] = 0$.
- $\mathrm{E}\left[Z_i X_i^\top\right]$ has rank k.
- $A_n \to_p A$, where A has rank $l \geq k$.
- $\mathrm{E}[X_{i,j}^2] < \infty$ for all $j = 1, \ldots, k$.
- $\mathrm{E}[Z_{i,j}^2] < \infty$ for all $j = 1, \ldots, l$.

Write

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

The last two assumptions imply that

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

By the WLLN,

$$n^{-1} \sum_{i=1}^n X_i Z_i^\top \to_p \mathrm{E}[X_i Z_i^\top].$$

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

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

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

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

Next, by the WLLN,

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

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

To show asymptotic normality, we will need the following three assumptions in addition to the above.

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

Write

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

The last two assumptions imply that the variance of $Z_i U_i$, $\mathrm{E}[U_i^2 Z_i Z_i^\top]$, is finite. By the CLT,

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

Define

$$\begin{aligned} Q &= \mathrm{E}[Z_i X_i^\top], \\ \Omega &= \mathrm{E}[U_i^2 Z_i Z_i^\top]. \end{aligned}$$

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:

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

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

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

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

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 provide heuristic arguments for consistency and asymptotic normality.

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

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

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

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

The formal proof of consistency requires regularity conditions: (i) compactness of B, (ii) continuity of $g(w, b)$ in b for each w, (iii) *uniform* convergence of the criterion over $b \in B$ in (3), and (iv) unique identification, that is, β is the unique minimizer of the population criterion.

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

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

In fact, it is sufficient if $\widehat{\beta}_n^{GMM}$ solves the first-order conditions approximately: the right-hand side of (4) may be $o_p(n^{-1/2})$ rather than exactly zero. This accommodates numerical optimization where the solution is found up to a tolerance that shrinks faster than $n^{-1/2}$.

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

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

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

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

$$+ \left(n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{GMM} \right)}{\partial b^\top} \right)^\top A_n^\top A_n \left(n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{*} \right)}{\partial b^\top} \right)$$

We can write

$$n^{1/2} \left(\widehat{\beta}_n \right.$$

$$= - \left(\left(n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{GMM} \right)}{\partial b^\top} \right)^\top A_n^\top A_n \left(n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{*} \right)}{\partial b^\top} \right) \right)^{-1}$$

$$\times \left(n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{GMM} \right)}{\partial b^\top} \right)^\top A_n^\top A_n n^{-1/2} \sum_{i=1}^{n}$$

Since $\mathrm{E}[g(W_i, \beta)] = 0$, under regularity conditions,

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

(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, under regularity conditions,

$$n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{GMM} \right)}{\partial b^\top} \to_p \mathrm{E}\left[\frac{\partial g\left(W_i, \beta\right)}{\partial b^\top} \right],$$

$$n^{-1} \sum_{i=1}^{n} \frac{\partial g \left(W_i, \widehat{\beta}_n^{*} \right)}{\partial b^\top} \to_p \mathrm{E}\left[\frac{\partial g\left(W_i, \beta\right)}{\partial b^\top} \right],$$

and that the matrix

$$\left(\mathrm{E}\left[\frac{\partial g\left(W_i, \beta\right)}{\partial b^\top} \right] \right)^\top A^\top A \left(\mathrm{E}\left[\frac{\partial g\left(W_i, \beta\right)}{\partial b^\top} \right] \right)$$

is invertible. Then,

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

where

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

$$Q = \mathrm{E}\left[\frac{\partial g\left(W_i, \beta\right)}{\partial b^\top} \right],$$

$$\Omega = \mathrm{E}[g\left(W_i, \beta\right) g\left(W_i, \beta\right)^\top].$$

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

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