Economics 522A, Spring 2007

Lecture Note 23: Generalized Method of Moments (Corrected 4/25/07)

(Reading: Ruud Ch. 21)

1 GMM Problem and Examples

The generalized method of moments (GMM) framework generalizes the linear IV model, to other settings involving orthogonality or other moment conditions.

We have observations $w_1, \ldots, w_n$, IID from an unknown distribution. Typically, w_i will be a vector, containing all variables observed in our data set.

Suppose we have a "moment function" $g(w, \theta)$ such that, at the "true" value of some parameter vector θ,

$$E[g(w_i, \theta)] = 0$$

Here $g(\cdot, \cdot)$ is a m-vector valued function so that the preceding display should be interpreted as m equality restrictions. We assume that the parameter of interest θ is $k \times 1$, where $k \leq m$.

To clarify the notation, let us use θ_0 to denote the true value of θ. So $g(w_i, \theta)$ is the function evaluated at some arbitrary value of θ, whereas $g(w_i, \theta_0)$ is the function evaluated at the true value of θ.

Let B_n be a $m \times m$ positive semidefinite matrix. So for a m-vector v,

$$(v' B_n v)^{1/2}$$

can be thought of as a "distance" of the vector v from 0. The GMM estimator is the solution to the following minimization problem:

$$\min_{\theta} \hat{Q}_n(\theta) = \left[\frac{1}{n} \sum_{i=1}^{n} g(w_i, \theta) \right]' B_n \left[\frac{1}{n} \sum_{i=1}^{n} g(w_i, \theta) \right].$$

The interpretation is that $\hat{\theta}$ tries to minimize the distance of

$$\frac{1}{n} \sum_{i=1}^{n} g(w_i, \theta)$$

from 0; that is, the GMM estimator tries to set the sample version of $E[g(w_i, \theta)]$ as close as possible to 0.

Example: Linear IV

Suppose $w_i = (y_i, x_i, z_i)$, where

$$y_i = x_i' \theta_0 + u_i,$$

but x_i may be correlated with u_i. The variable z_i is an instrumental variable:

$$E[z_i u_i] = 0.$$

Rewrite this as

$$E[z_i(y_i - x_i'\theta_0)] = 0.$$

So we can take

$$g(w_i, \theta) = z_i(y_i - x_i'\theta).$$

Example: Euler Equations (Hansen and Singleton, 1982)

Suppose a consumer is choosing a consumption stream $c_1, \ldots, c_T$ to maximize expected utility, under a constant relative risk aversion utility function

$$u(c) = \frac{c^\gamma - 1}{\gamma}.$$

The consumer's maximization problem is

$$\max_{c_1,\ldots,c_T} E\left[\sum_{t=1}^{T} \beta^{t-1} u(c_t)\right]$$

subject to a dynamic budget constraint. Here β is the rate of time preference.

The firsr-oder (Euler) conditions for a maximum are

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1} - 1|\mathcal{I}_t\right] = 0,$$

for all t, where $\mathcal{I}_t$ denotes the information available to the agent at time t.

Let $\theta = (\beta, \gamma)$ and

$$\rho(w_t, \theta) = \beta\left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1} - 1.$$

So we have

$$E[\rho(w_t, \theta_0)|\mathcal{I}_t] = 0.$$

Now, suppose x_t are variables that are included in the information set at time t. For example, it could include lagged values of the consumption variable and other lagged variables such as income. Then

$$E[x_t\rho(w_t, \theta_0)] = 0.$$

So we can define

$$g(w_t, \theta) = x_t\rho(z_t, \theta),$$

and we have that

$$E[g(w_t, \theta_0)] = 0.$$

This is a nonlinear version of the previous instrumental variables problem.

Example: MLE

The maximum likelihood estimator (and conditional maximum likelihood estimator) can sometimes be viewed as a GMM estimator.

Suppose that we have data w_i IID from a distribution with density $f(w_i;\theta_0)$. Recall that

$$Q(\theta) := E\left[\log f(w_i;\theta)\right]$$

is maximized at the true value θ_0. Suppose that $\log f(w_i;\theta)$ is continuously differentiable in θ and that we can interchange the order of differentiation and integration. Then, at the true value, we will have

$$E\left[\frac{\partial \log f(w_i;\theta_0)}{\partial\theta}\right] = 0.$$

So we can take

$$g(w_i,\theta) = \frac{\partial \log f(w_i;\theta)}{\partial\theta}.$$

In this case, g is the score function and has the same dimension as θ.

2 Consistency of GMM Estimator

We will only give a heuristic discussion of consistency and asymptotic normality of the GMM estimator. For a thorough treatment, see: Newey and McFadden (1994).

Let

$$\hat{g}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta),$$

$$g_0(\theta) = E[g(w_i,\theta)] \qquad (= 0 \text{ at } \theta = \theta_0).$$

Define

$$\hat{Q}_n(\theta) = \hat{g}_n(\theta)' B_n \hat{g}_n(\theta).$$

The first step to showing consistency is to show that the the function $\hat{Q}_n$ converges "uniformly" to a limiting objective function.

Assume that $B_n \xrightarrow{\text{p}} B$, where B is positive semidefinite and finite. Define

$$Q(\theta) = g_0(\theta)' B g_0(\theta).$$

This is a continuous function of θ, since g_0 is continuous and W is positive semidefinite.

Then under conditions on the function g, it can be shown that

$$\sup_{\theta\in\Theta} |\hat{Q}_n(\theta) - Q(\theta)| \xrightarrow{\text{p}} 0.$$

Heuristically, this strong form of convergence will imply that the minimizer of $\hat{Q}_n(\theta)$ will converge to the minimizer of $Q(\theta)$. Recall that by assumption,

$$g_0(\theta_0) = E[g(z_i,\theta_0)] = 0.$$

Suppose we can also show that

$$B g_0(\theta) \neq 0 \quad \forall \theta \neq \theta_0.$$

Let $R'R = B$. Then the previous display implies

$$Rg_0(\theta) \neq 0 \quad \forall \theta \neq \theta_0.$$

Therefore

$$Q(\theta) = (Rg_0(\theta))'(Rg_0(\theta)) > 0 \quad \forall \theta \neq \theta_0.$$

So $Q(\theta)$ is uniquely minimized at $\theta = \theta_0$.

If these conditions are satisfied, then we will have that $\hat{\theta} \xrightarrow{\text{p}} \theta_0$.

3 Asymptotic Normality of GMM Estimator

In the GMM problem, the first order condition for a minimum is

$$\left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(w_i, \hat{\theta})\right]' B_n \left[\frac{1}{n}\sum_{i=1}^{n} g(w_i, \hat{\theta})\right] = 0.$$

Here, we are using ∇_θ to denote the gradient.

Expand the function $g(w, \theta)$:

$$g(w_i, \hat{\theta}) = g(w_i, \theta_0) + \nabla_\theta g(w_i, \bar{\theta})(\hat{\theta} - \theta_0).$$

Substitute this into the first order condition:

$$\left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(w_i, \hat{\theta})\right]' B_n \left[\frac{1}{n}\sum_{i=1}^{n} g(w_i, \theta_0) + \left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(w_i, \bar{\theta})\right](\hat{\theta} - \theta_0)\right] = 0.$$

Let

$$\hat{G}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g(w_i, \theta).$$

Rearrange the first order condition to get

$$\sqrt{n}(\hat{\theta} - \theta_0) = -(\hat{G}_n(\hat{\theta})' B_n \hat{G}_n(\bar{\theta}))^{-1} \hat{G}_n(\hat{\theta})' B_n \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_0).$$

Then, under suitable conditions to apply the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(w_i, \theta_0) \xrightarrow{\text{d}} N(0, \Omega),$$

$$B_n \xrightarrow{\text{p}} B,$$
$$\hat{G}_n(\hat{\theta}) \xrightarrow{\text{p}} G,$$
$$\hat{G}_n(\bar{\theta}) \xrightarrow{\text{p}} G.$$

So

$$(\hat{G}_n(\hat{\theta})'B_n\hat{G}_n(\bar{\theta}))^{-1}\hat{G}_n(\hat{\theta})'B_n \xrightarrow{\text{p}} (G'BG)^{-1}G'B,$$

and by the Slutsky lemma,

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\text{d}} N(0,(G'BG)^{-1}G'B\Omega BG(G'BG)^{-1}).$$

Optimal weighting matrix: Suppose $B=\Omega^{-1}$. (That is, $B_n \xrightarrow{\text{p}} \Omega^{-1}$.). Then

$$V=(G'BG)^{-1}G'B\Omega BG(G'BG)^{-1}=(G'\Omega^{-1}G)^{-1}.$$

This can be shown to be the best variance, in the sense that the difference between any other feasible V and this variance is a positive semidefinite matrix.

4 Estimating the Variance of GMM Estimator

Recall the asymptotic variance of GMM is

$$(G'BG)^{-1}G'B\Omega BG(G'BG)^{-1},$$

where

$$G=E[\nabla_\theta g(w_i,\theta_0)],$$
$$B=\text{plim } B_n,$$
$$\Omega=E[g(w_i,\theta_0)g(w_i,\theta_0)'].$$

We can estimate

$$\hat{\Omega}=\frac{1}{n}\sum_{i=1}^n g(w_i,\hat{\theta})g(w_i,\hat{\theta})',$$

$$\hat{G}=\frac{1}{n}\sum_{i=1}^n \nabla_\theta g(w_i,\hat{\theta}).$$

So the estimated variance is

$$\hat{V}=(\hat{G}'B_n\hat{G})^{-1}\hat{G}'B_n\hat{\Omega}B_n\hat{G}(\hat{G}'B_n\hat{G})^{-1}.$$

Optimal GMM: Recall that the ideal choice of the weight matrix is

$$B_n \xrightarrow{\text{p}} \Omega^{-1}=\left(E[g(w_i,\theta_0)g(w_i,\theta_0)']\right)^{-1}.$$

But since θ_0 is not known, how do we implement this?

One solution (Hansen, 1982):

First, use a suboptimal $B_n^{(1)}$, solve the GMM problem, to get an initial estimator $\hat{\theta}_1 \xrightarrow{\text{p}} \theta_0$.

Calculate

$$\hat{\Omega}_1 = \frac{1}{n} \sum_{i=1}^{n} g(w_i, \hat{\theta}_1) g(w_i, \hat{\theta}_1)',$$

$$B_n^{(2)} = (\hat{\Omega}_1)^{-1}.$$

Then re-estimate θ using the new weight matrix $B_n^{(2)}$:

$$\hat{\theta}_2 = \arg\min_{\theta \in \Theta} \left[\frac{1}{n} \sum_{i=1}^{n} g(w_i, \theta) \right]' B_n^{(2)} \left[\frac{1}{n} \sum_{i=1}^{n} g(w_i, \theta) \right].$$

Since $B_n^{(2)} \xrightarrow{\text{p}} \Omega^{-1}$, the final estimator $\hat{\theta}_2$ will have asymptotic variance $(G'\Omega^{-1}G)^{-1}$.

5 References

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50(4), 1029-54.

Hansen, L., and Singleton, K., 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50(5), 1269-86.

Newey, W., and McFadden, D., 1994, "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden, (eds.) *Handbook of Econometrics*, Vol IV, 2111-2245, North Holland: Amsterdam.