14.384 Time Series Analysis
Fall 2008

14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 8, 2007

Lecture 18

GMM

A dynamic stochastic general equilibrium (DSGE) model can be estimated in three ways:

1. GMM
2. MLE
3. Bayesian

Today we will talk about GMM

GMM

We have data z_t, parameter θ, and moment condition $Eg(z_t, \theta_0) = 0$. This moment condition identifies θ. Many problems can be formulated in this way.

Examples

- **OLS**: we have $E(y_t|x_t) = x_t'\beta$. We can write this as a conditional (on x) moment condition, $E(y_t - x_t'\beta|x_t) = 0$ or an undconditional moment condition, $E[x_t(y_t - x_t'\beta)] = 0$
- **IV**:

$$\begin{aligned} y_t =& x_t'\beta + e_t \\ E(e_t|z_t) =& 0 \end{aligned}$$

  This gives the moment condition $E[z_t(y_t - x_t'\beta)] = 0$
- **Euler Equation**: This was the application for which Hansen developed GMM. Suppose we have CRRA utility, $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$. The first order condition from utility maximization gives us the Euler equation,

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} R_{t+1} - 1 \right] = 0$$

  This moment condition can be used to estimate γ and β.

Estimation

Take the moment condition and replace it with its sample analog:

$$Eg(z_t, \theta) \approx \frac{1}{T} \sum_{t=1}^{T} g(z_t, \theta) = g_T(\theta)$$

Our estimate, $\hat{\theta}$ will be such that $g_T(\hat{\theta}) \approx 0$. Let's suppose θ is $k \times 1$ and $g_T(\theta)$ is $n \times 1$. If $n < k$, then θ is not identified. If $n = k$ (and $\frac{dg}{d\theta}(\theta_0)$ is full rank), then we are just identified, and we can find $\hat{\theta}$ such that $g_T(\hat{\theta}) = 0$. If $n > k$ (and the rank of $g'(\theta_0) > k$), then we are overidentified. In this case, it will generally be impossible to find $\hat{\theta}$ such that $g_T(\hat{\theta}) = 0$, so we instead minimize a quadratic form

$$\hat{\theta} = \arg\min_{\theta} g_T(\theta)' W_T g_T(\theta)$$

where W_T is symmetric and positive definite.

Some things we want to know:

1. asymptotics of $\hat{\theta}$
2. efficient choice of W_t
3. test of overidentifying restrictions

To analyze the asymptotics, we need some assumptions.

Assumptions

1. $Eg(\theta_0) = 0$ only at θ_0 (identification)
2. $\frac{\partial}{\partial\theta} Eg(\theta)|_{\theta=\theta_0} = R$ is bounded and continuous
3. $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \theta_0) \Rightarrow N(0, S)$
4. $\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g}{\partial\theta}(z_t, \theta) \xrightarrow{p} R(\theta)$
5. $W_T \xrightarrow{p} W$

Note that we're just assuming that some CLT and LLN apply to the above quantities. We can deal with non-iid observations as long as a CLT and LLN apply.

Asymptotic Distribution

Theorem 1. *Under these assumptions, $\hat{\theta}$ is consistent and asymptotically normal.*

$$\sqrt{T}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma)$$
$$\Sigma = (RWR')^{-1}(RWSW'R')(RWR')^{-1}$$

Proof. To prove this, we take a Taylor expansion of the first order condition. The first order condition is:

$$g_T(\theta)' W_T \frac{\partial g_T(\theta)}{\partial\theta}|_{\theta=\hat{\theta}} = 0$$

Expanding $g_T(\theta)$ around θ_0

$$\sqrt{T} g_T(\hat{\theta}) = \sqrt{T} g_T(\theta_0) + \frac{\partial g(\theta_0)}{\partial\theta} \sqrt{T}(\hat{\theta} - \theta_0) + o_p(1)$$

We've assumed that $\sqrt{T} g_T(\theta_0) \Rightarrow N(0, S)$, $\frac{\partial g(\theta_0)}{\partial\theta} \xrightarrow{p} R$. Then, putting our expansion into the first order condition, rearranging, and using these convergence results, we get

$$\sqrt{T}(\hat{\theta} - \theta_0) = (\frac{\partial g_T}{\partial\theta}' W_T \frac{\partial g_T}{\partial\theta})^{-1} (\frac{\partial g_T}{\partial\theta}' W_T \sqrt{T} g_T)$$
$$\Rightarrow (RWR')^{-1} RWN(0, S)$$

□

Efficient Weighting Matrix

Lemma 2. *The efficient choice of W is S^{-1}. This choice of W gives asymptotic variance $\tilde{\Sigma} = (RS^{-1}R)^{-1}$. For any other W, $\Sigma - \tilde{\Sigma}$ is positive semi-definite.*

In practice, we do not know S.

Definition 3. *Feasible efficient GMM*

- Choose arbitrary W, say I, get initial estimate θ_1, use it to calculate $\hat{S}$ (in time series, we need to use Newey-West to estimate $\hat{S}$)
- Re-run GMM with $W_T = \hat{S}^{-1}$

We can also iterate this procedure, to do *continuous updating estimator* (CUE) GMM. However, CUE is no more efficient than two-step GMM. In IV, CUE is the same as LIML.

Tests of Overidentifying Restrictions

Definition 4. *J test*: Under the null that all moment conditions are true, the statistic, J, converges to a χ^2_{n-k}

$$J = [\sqrt{T}g(z,\hat{\theta})]'\hat{\Sigma}^{-1}[\sqrt{T}g(z,\hat{\theta})] \Rightarrow \chi^2_{n-k}$$

Remark 5. This is a test of all the restrictions jointly. If we reject, it does not tell which moment condition is wrong. It merely means the moment conditions contradict one another.

Simulated GMM

Suppose our moment conditions are too complicated to calculate directly, but we can simulate data given our parameters and we can compute the moment condition from observed data. For example, suppose we have a model that we think should explain some moments of the data, $h(z_t)$. We want to choose the parameters of our model, θ, such that $Eh(z_t) = \phi(\theta)$, but $\phi()$ is too hard to calculate. Instead, we can simulate data given the parameters, $y_j(\theta)$. Then, we choose θ such that

$$E(h(z_t) - Eh(y_j(\theta_0))) = 0$$

This idea is broadly applicable. We will follow on the approach of Lee & Ingram (1991). They focus on time series applications. Another example is Pakes & Pollard, or McFadden, they use simulated GMM for models with discrete outcomes. In this case, $h()$ might not be continuous, so there are some additional technical issues to deal with.

Setup Let

- $H_T(z) = \frac{1}{T}\sum_{t=1}^{T} h(z_t)$ be the average of the observed moments.
- $H_N(y(\theta)) = \frac{1}{N}\sum_{j=1}^{N} h(y_j(\theta))$ be the average of the simulated moments given θ

Our moment function is

$$g_T(\theta) = H_T(z) - H_N(y(\theta))$$

and the objective function is

$$\min_{\theta} g_T(\theta)'W_T g_T(\theta)$$

Assume that $h(y_j(\theta))$ is continuous in θ, *ie*

$$\lim_{\delta\to 0} E\left[\sup_{|\theta_1-\theta_2|<\delta} |h(y_j(\theta_1)) - h(y_j(\theta_2))|\right] = 0$$

Remark 6. In practice, this condition requires that each time you simulate $y_j(\theta)$, you should use the same underlying error terms, *ie* you should reset the seed of your random number generator to the same value (and make sure that $h(y_j())$ is continous).

What conditions do we need on this setup to satisfy the GMM assumptions above?

- $H_T(z) \to^{a.s} \phi(\theta_0)$ and $H_N(z) \to \phi(\theta)$.
- $\{z_t\}$ & $\{y_j(\theta)\}$ are independent
- We need a central limit theorem for both $H_T(z)$ and $H_N(y(\theta))$

$$\sqrt{T}(H_T(z) - \phi(\theta_0) \Rightarrow N(0, S)$$

$$\sqrt{N}(H_N(y(\theta_0)) - \phi(\theta_0) \Rightarrow N(0, S)$$

  Note that this will require that both z and $y(\theta_0)$ are stationary. For z, we can just assume this. However, for $y(\theta_0)$, it is a bit tricky. In time series, for $y(\theta)$ to be stationary, we need make sure that its initial value is drawn from its stationary distirbution

- The asymptotic variance will have an additional term due to the variation in simulation, $\Sigma_{SMM} = (1 + \frac{T}{N})\Sigma$