14.384 Time Series Analysis
Fall 2008

14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 13, 2007

Lecture 19

# Simulated MM and Indirect Inference

## Indirect Inference

Suppose we are interested in parameter $\beta \in C$ with true value $\beta_0$. We observe data $\{x_t\}_{t=1}^T$. We have a model that we can simulate to generate $\{y_j(\beta)\}_{j=1}^S$. We don't necessarily believe that our model is the true DGP, but we do think our model can explain some features of the data. These features we want to explain can be written as a function $\theta(\{x_t\})$. Let

$$\hat{\theta}_T = \theta(\{x_t\})$$
$$\hat{\theta}_s^\beta = \theta(\{y_j(\beta)\})$$

We'll assume that $\theta()$ is an extremum estimator.

$$\hat{\theta}_T = \theta(\{x_t\}) = \arg\max_\theta Q_T(\{x_t\}, \theta)$$
$$\hat{\theta}_s^\beta = \theta(\{y_j(\beta)\}) = \arg\max_\theta Q_S(\{y_j(\beta)\}, \theta)$$

For example, $\theta$ could be simple sample means or moments, or regression coefficients, or more generally parameters from some sort of auxillary model. We estimate $\beta$ by matching $\hat{\theta}_T$ to $\hat{\theta}_S^\beta$

$$\hat{\beta} = \arg\min_\beta (\hat{\theta}_T - \hat{\theta}_S^\beta)' W_T (\hat{\theta}_T - \hat{\theta}_S^\beta)$$

This estimator is discussed by Smith (1993) for the case where $Q_T(\cdot, \theta)$ is a pseudo-loglikelihood. Gourierox, Monfort, and Renault (1993) consider a more general setup. We will go through Smith's setup. $Q_T(\cdot, \theta)$ is a pseudo-loglikelihood:

$$Q_T(\{x_t\}, \theta) = \sum_{t=p}^T \log f(x_t, ..., x_{t-p}; \theta)$$

We call this a pseudo-loglikelihood because we can allow $f$ to be misspecified, ie $f$ need not be the ture density of $x_t$. Assume

1. $x_t$ and $y_t(\beta)$ $\forall \beta$ are stationary and ergodic
2. $\exists$ a unique $\beta_0$ such that $(x_t, ..., x_{t+p}) =^d (y_s(\beta_0), ..., y_{s+p}(\beta_0))$ so $\theta(x_t) = \theta(y_s(\beta_0)) = \theta_0$
3. $f$ is well-behaved (has several continuous, well-bounded derivatives)
4. $\arg\max_\theta E \log f(y_s(\beta), ..., y_{s-p}(\beta); \theta) = \theta^\beta = h(\beta)$ and $\theta^\beta$ is the unique maximizer for each $\beta$

Under these assumptions, $\hat{\theta}_s^\beta \to \theta^\beta = h(\beta)$ and $\hat{\theta}_T \to \theta_0 = h(\beta_0)$

For the asymptotic distribution, we need some notation. Let:

$$A_\beta(\theta) = E\nabla^2 \log f(\{y(\beta)\},\theta)$$
$$B_\beta(\theta) = \Gamma_0^\beta(\theta) + \sum_{k=1}^{\infty}(\Gamma_k^\beta(\theta) + \Gamma_{-k}^\beta(\theta))$$

where $\Gamma_k^\beta(\theta) = \text{cov}(\nabla \log f(y_s(\beta),\theta), \log f(y_{s-k}(\beta),\theta))$. Assume a CLT:

$$\frac{1}{\sqrt{T}}\nabla Q_s(\{y(\beta)\},\theta) \Rightarrow N(0, B_\beta(\theta))$$

then

$$\sqrt{S}(\hat{\theta}_s^\beta - \theta_\beta) \Rightarrow N(0, A_\beta^{-1} B_\beta A_\beta^{-1})$$
$$\sqrt{T}(\hat{\theta}_T - \theta_0) \Rightarrow N(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1})$$

We get this sandwhich form for the variance because we have not assumed that the likelihood is correctly specified. If the likelihood ($f$) is correctly specified, then the information equality would hold and $A = B$. These results follow from the general theory of extremum estimators.

## Extremum Estimators

Extremum estimators are very common in econometrics.

$$\hat{\theta} = \arg\min Q_T(\{x_t\},\theta)$$

Most estimators, such as least squares, GMM, and MLE, fit into this framework. Under some regularity conditions (including $Q$ is differentiable and $\nabla Q$ satisfies a CLT), if $Q_T(\{x_t\},\theta) \rightarrow_{a.s.} Q_\infty(\theta)$ uniformly and

$$\theta_0 = \arg\min Q_\infty(\theta)$$

then $\hat{\theta} \xrightarrow{p} \theta_0$ and $\hat{\theta}$ is asymptotically normal. As with GMM, we can show this by taking a Taylor expansion of the first order condition:

$$0 = \nabla Q_T(\{x_T\},\hat{\theta})$$
$$= \nabla Q_T(\{x_T\},\theta_0) + (\hat{\theta} - \theta_0)\nabla^2 Q_T(\{x_T\},\theta_0) + o_p$$
$$\sqrt{T}(\hat{\theta} - \theta_0) = \frac{\frac{1}{\sqrt{T}}\nabla Q_T(\{x_T\},\theta_0)}{\frac{1}{T}\nabla^2 Q_T(\{x_T\},\theta_0)}$$
$$\Rightarrow N(0, A^{-1}BA^{-1})$$

## More Indirect Inference

So far, we have shown that $\hat{\theta}_T$ and $\hat{\theta}_s^\beta$ are consistent and asymptotically normal. Now, we want to show that $\hat{\beta}$ is consistent and asymptotically normal. We need some additional assumptions:

1. $h_s(\beta)$ and $h(\beta)$ are continuous and several times differentiable. Let $J(\beta) = \nabla h(\beta)$.
2. $\nabla h_s(\beta) \xrightarrow{p} \nabla h(\beta)$ uniformly in $\beta$
3. $S = \tau T$

Then $\hat{\beta} \xrightarrow{p} \beta$ and

$$\sqrt{T}(\hat{\beta} \xrightarrow{p} \beta) \Rightarrow N\left(0, (1+\frac{1}{\tau})\Sigma(\beta_0)\right)$$

where $\Sigma(\beta_0) = K^{-1}J'W\Omega(\theta_0)WJK^{-1}$, $\Omega(\theta_0) = A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}$ and $K(\beta_0) = J(\beta_0)'WJ(\beta_0)$.

*Proof.* As usual, this is just sketch and not fully rigorous. Consider the FOC for

$$\begin{aligned}\hat{\beta} =& \arg\min_{\beta}(\hat{\theta}_T - \hat{\theta}_S^{\beta})'W_T(\hat{\theta}_T - \hat{\theta}_S^{\beta}) \\ =& \arg\min_{\beta}(\hat{\theta}_T - h_S(\beta))'W_T(\hat{\theta}_T - h_S(\beta))\end{aligned}$$

We have:

$$0 = \frac{\partial}{\partial\beta}h_s(\hat{\beta})'W_t(\hat{\theta}_T - h_s(\hat{\beta}))$$

Taking a Taylor expansion of $h_s(\hat{\beta}) \approx h_s(\beta_0) + \nabla h_s(\beta_0)(\hat{\beta} - \beta_0)$, taking the limit, and multiplying by $\sqrt{T}$:

$$\begin{aligned}0 =& J'W\sqrt{T}(\hat{\theta}_T - h_s(\beta_0)) + J'WJ\sqrt{T}(\hat{\beta} - \beta_0) \\ J'WJ\sqrt{T}(\hat{\beta} - \beta_0) =& J'W\sqrt{T}(\hat{\theta}_T - h_s(\beta_0)) \\ =& J'W(\hat{\theta}_T - \theta_0 + \theta_0 - h_s(\beta_0))J'WJ\sqrt{T}(\hat{\beta} - \beta_0) \\ \Rightarrow& J'W(N(0,\Omega) + \frac{1}{\sqrt{\tau}}N(0,\Omega))\end{aligned}$$

□

Also, as usual, the efficient choice of $W = \Omega^{-1}$ with a feasible estimate $W_T^* = \hat{\Omega}^{-1}$ where $\hat{\Omega}$ is formed from some consistent estimate of $\theta_0$.

## Test of Overidentifying Restrictions

If $\dim(\theta) = n > \dim(\beta) = k$, then we can test the overidentifying restrictions with

$$\begin{aligned}\mathcal{Z}_t =& T\frac{1}{1+\frac{1}{\tau}}(\hat{\theta}_T - h_s(\hat{\beta}_T))'W_T(\hat{\theta}_T - h_s(\hat{\beta}_T)) \\ =& T\frac{1}{1+\frac{1}{\tau}}(\hat{\theta}_T - \hat{\theta}_s^{\beta})'W_T(\hat{\theta}_T - \hat{\theta}_s^{\beta}) \\ \Rightarrow& \chi^2_{n-k}\end{aligned}$$

## Example

Suppose we have a DSGE model where someone is maximizing something. For example

$$\begin{aligned}&\max E_0 \sum \omega^t \frac{c^{-\gamma} - 1}{\gamma} \\ &\text{s.t. } c_t + i_t = A\lambda_t k_t^{\alpha} \\ &\quad k_{t+1} = (1-\delta)k_t + i_t \\ &\quad \lambda_t = \rho\lambda_{t-1} + \epsilon_t \;\; \epsilon_t \sim N(0, \sigma^2)\end{aligned}$$

This model has many parameters $(\omega, \gamma, A, \alpha, \delta, \rho, \sigma^2)$ and would be very difficult to write down a likelihood or moment functions. Moreover, we don't really believe that this model is the true DGP and we don't want to use it to explain all aspects of the data. Instead we just want the model to explain some feature of the data, say the dynamics as captured by VAR coefficients. Also, although it is hard to write the likelihood function for this model, it is fairly easy to simulate the model. The we can use indirect inference as follows:

1. Estimate (possibly misspecified) VAR from data
2. Given $\beta$, simulate model, estimate VAR from simulations, repeat until minimize objective function

## Calibration

This was the method for evaluating models proposed by Kydland and Prescott. It has been very widely used. The steps are

1. Ask a question: either an assessment of a theoretical implication of a policy or testing the ability of a model to mimic features of actual data
2. Pick a class of models
3. Write down the equations
4. Your model should have implications to some questions with known answers (eg the labor share is 70%). You calibrate your parameters to match these answers. Kydland and Prescott describe this step as follows:

   > Some economic questions have known answers, the model should give approximately correct answers to them ... Data are used to calibrate the model economy ... to mimic the world as closely as possible along a limited number of dimensions. Calibration is not an attempt to assess a size of something; it's not estimation ... The parameter values selected are not the ones that provide the best fit in some statistical sense. (Kydland and Prescott)

5. Run an experiment to see if you model matches other aspects of the data.

This method was meant to get away from the way that statisticians put too much structure on the data. People criticize this method as not being rigorous and giving difficult to communicate results. There is no formal criteria for whether the model matches well in step 5. The judgement of good and bad fit is subjective. Hansen and Heckman argued that calibration can be thought of an informal version of estimation, so we might as well use the formal theory so that we have standard errors and formal testing. In fact when all of our calibration is take from moments in the data, calibration is the same a just-identified simulated GMM. If the calibration involve estimates taken from other studies, we can still use the simulated GMM framework, but we need to take into account the standard errors of the parameters taken from other studies.