Computational Economics and Econometrics Efficient Method of Moments

by

A. Ronald Gallant
Penn State University

Review of Ideas

- The ideas have already been presented in the habit model case study.

- EMM is a simulated method of moments estimator with an optimal choice of moments: the SNP scores.

  – Gallant, A. Ronald, and George Tauchen (1996), "Which Moments to Match?," *Econometric Theory* 12, 657–681.

  – Gallant, A. Ronald, and Jonathan R. Long (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika* 84, 125–141.

Review of Ideas

- The case study code was the preliminary version of the EMM code.
  - EMM code has a better user interface.
  - SNP scores are automatically generated from an SNP parmfile.
  - Efficiency improvements and parallelization.

- Change of notation to be consistent with SNP and EMM User's Guides: $\theta \to \rho$

Characteristics of Models of Specific Interest

- Likelihood not available.

- Prior information $\pi_1(\rho)$ on model parameters may be available.

- Prior information $\pi_2(\rho, \psi)$ on functionals of the model may be available, i.e. $\psi = \Psi(\mathcal{M}_\rho)$.

- Model can be simulated.

Example

Stocastic volatility model

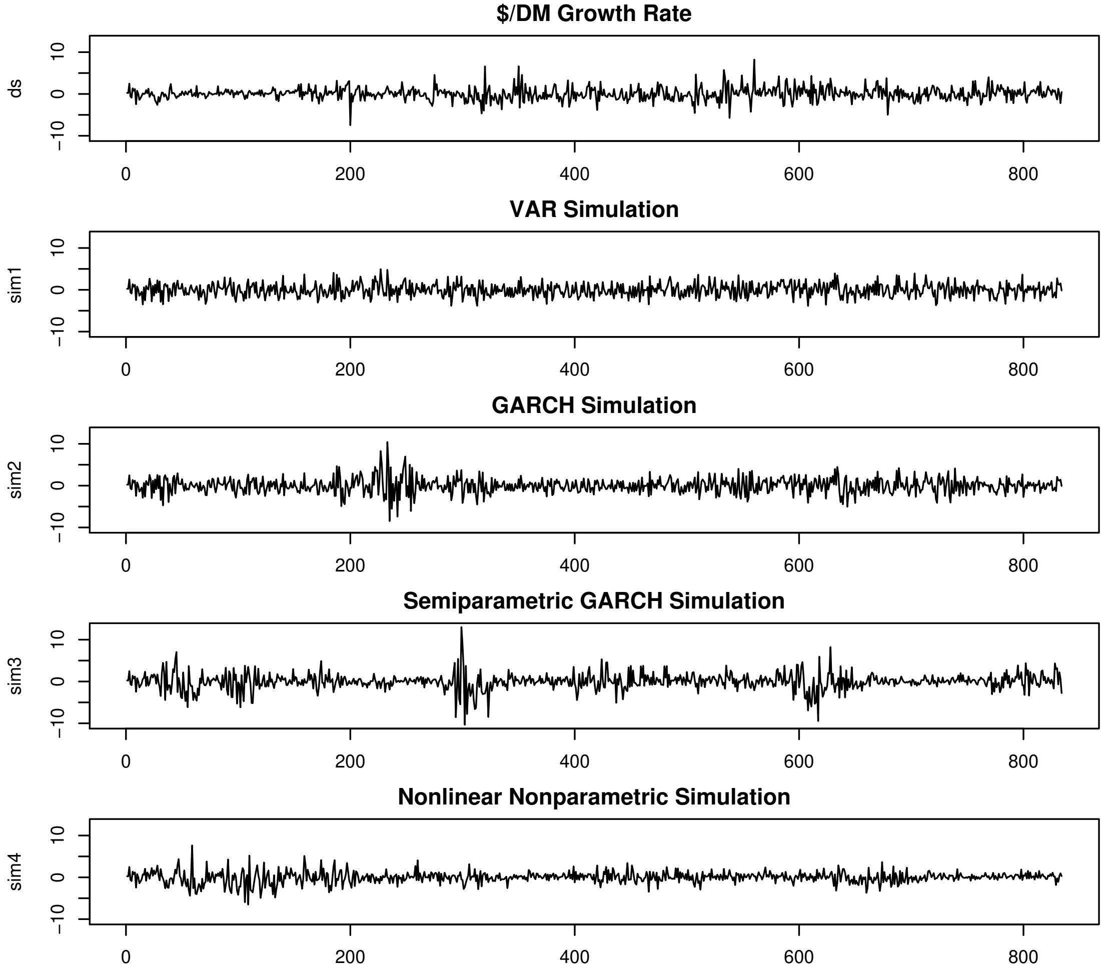
$$
\begin{aligned}
y_t &= a_0 + a_1(y_{t-1} - a_0) + \exp(v_t)u_{1t} \\
v_t &= b_0 + b_1(v_{t-1} - b_0) + u2_t \\
u_{1t} &= z_{1t} \\
u_{2t} &= s\left(r\, z_{1t} + \sqrt{1 - r^2}\, z_{2t}\right),
\end{aligned}
$$

where the errors $z_t = (z_{1t}, z_{2t})$ are iid $N_2(0, I)$.

These equations imply

$$
\mathrm{Var}(u) = \mathcal{E}(uu') = \begin{pmatrix} 1 & sr \\ sr & s^2 \end{pmatrix}
$$

Fig 1. Changes in Weekly \$/DM Exchange Rates: 1975–1990



The first panel is a plot of the data. The second is a simulation from an SNP fit with $(L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_x) = (1, 0, 0, 1, 0, 0, 0, 0)$; the third with (1,1,1,1,0,0,0,0), the fourth with (1,1,1,1,4,0,0,0), and the fifth with (1,1,1,1,4,0,1,0). L_v and L_w are set to zero; I_z, maxI_z, and I_x have no effect when $M = 1$; max$K_z = K_z$.

SMM with a GMM Criterion

- Output and parameters of the stochastic volatility model are

$$y_t \in \Re^1$$

$$\rho = (a_0, a_1, b_0, b_1, s, r) \in \Re^6$$

- Data are denoted as $\{\tilde{y}_t\}_{t=1}^n$, simulations as $\{\hat{y}_t\}_{t=1}^N$.

Some Notation

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$$

$$S_t = \left[\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} - \bar{y} \right] \left[\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} - \bar{y} \right]'$$

$$m_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \text{vech}(S_t) \end{pmatrix}$$

$\tilde{m}_t$ denotes evaluation at data

$\hat{m}_t$ denotes evaluation at a simulation

A GMM Criterion – Moment Functions

Moment function for data:

$$\tilde{m}_n = \frac{1}{n} \sum_{t=1}^{n} \tilde{m}_t$$

Moment function for a simulation:

$$\hat{m}_N(\rho) = \frac{1}{N} \sum_{t=1}^{N} \hat{m}_t$$

GMM Cross Sectional Weight Function

$\tilde{W}_n$ is an estimate of the variance of $\sqrt{n}\,\tilde{m}_n$

$$\tilde{W}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{m}_i - \tilde{m}_n\right)\left(\tilde{m}_i - \tilde{m}_n\right)'$$

GMM Time Series Weight Function

$\tilde{W}_n$ is an estimate of the variance of $\sqrt{n}\,\tilde{m}_n$

$$\tilde{W}_n = \sum_{\tau=-[n^{1/5}]}^{[n^{1/5}]} w\left(\frac{\tau}{[n^{1/5}]}\right)\tilde{W}_{n\tau}$$

where

$$w(u) = \begin{cases} 1 - 6|u|^2 + 6|u|^3 & \text{if } 0 < u < \frac{1}{2} \\ 2(1 - |u|)^3 & \text{if } \frac{1}{2} \le u < 1 \end{cases}$$

$$\tilde{W}_{n\tau} = \begin{cases} \frac{1}{n}\sum_{t=1+\tau}^{n}(\tilde{m}_t - \tilde{m}_n)(\tilde{m}_{t-\tau} - \tilde{m}_n)' & \tau \ge 0 \\ \tilde{W}'_{n,-\tau} & \tau < 0 \end{cases}$$

GMM Criterion Function

$$s_n(\rho) = \frac{1}{2}\left[\tilde{m}_n - \widehat{m}_N(\rho)\right]' \left(\tilde{W}_n\right)^{-1} \left[\tilde{m}_n - \widehat{m}_N(\rho)\right]$$

EMM Criterion Function – 1

EMM is SMM with a different GMM criterion function.

The GMM moment equations above

$$\tilde{m}_n - \hat{m}_N(\rho)$$

are replaced by the scores from an SNP fit

$$m(\rho, \tilde{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial}{\partial \theta} \log f(\hat{y}_t | \hat{x}_{t-1}, \tilde{\theta}),$$

where $f(y_t | x_{t-1}, \theta)$ is an SNP density chosen by BIC and with its parameters θ set to the value $\tilde{\theta}$ estimated from the data.

Remember that a tilde (˜) represents something computed from data and a circumflex (ˆ) represents something computed from a simulation with parameter values of the structural model set to ρ. **Details follow**.

EMM Criterion Function – 2

- For any QMLE estimator

$$\tilde{\theta} = \underset{\theta}{\text{argmax}} \frac{1}{n} \sum_{t=1}^{n} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \theta),$$

  a sample average satisfies

$$0 = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta})$$

  because these are the first order conditions of the optimization problem.

- The SNP estimator is a QMLE estimator.

- Therefore a large simulation from a correctly specified structural model $p(y_t | x_{t-1}, \rho)$ will satisfy

$$0 = m(\rho, \tilde{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial}{\partial \theta} \log f(\hat{y}_t | \hat{x}_{t-1}, \tilde{\theta}),$$

  except for sampling variation in $\tilde{\theta}$. The equality holds exactly in the limit as n and N tend to infinity.

EMM Criterion Function – 3

EMM finds ρ that satisfies the first order conditions $0 = m(\rho, \tilde{\theta})$ as nearly as possible by computing

$$\hat{\rho} = \underset{\rho}{\text{argmin}}\, m'(\rho, \tilde{\theta})(\tilde{W})^{-1} m(\rho, \tilde{\theta}),$$

where

$$m(\rho, \tilde{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial}{\partial \theta} \log f(\hat{y}_t | \hat{x}_{t-1}, \tilde{\theta})$$

$$\tilde{W} = \frac{1}{n} \sum_{t=1}^{n} [\frac{\partial}{\partial \theta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta})][\frac{\partial}{\partial \theta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta})]'$$

This estimator achieves the same efficiency as maximum likelihood:

Gallant, A. Ronald, and Jonathan R. Long (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika* 84, 125–141.

Asymptotics

Under weak regularity conditions that accommodate both time series and cross sectional data (Gallant, 1987) $\hat{\rho}_n$ tends to the parameter value ρ^o that minimizes

$$s^o(\rho) = \lim_{n\to\infty} s_n(\rho)$$

and $\sqrt{n}(\hat{\rho}_n - \rho^o)$ is asymptotically normal with mean zero and variance $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$, where $\mathcal{J}$ is the Hessian

$$\mathcal{J} = \frac{\partial}{\partial\rho\partial\rho'}\, s^o(\rho^o)$$

and $\mathcal{I}$ is Fisher's information

$$\mathcal{I} = \text{Var}\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right] = \mathcal{E}\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right]\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right]'$$

For SNP, $\mathcal{I} = \mathcal{J}$ so that only one of the two has to be computed; e.g. correctly specified mle or GMM with correct weight matrix.

Computations

For $\quad s_n(\rho) = \frac{1}{2}\left[\tilde{m}_n - \widehat{m}_N(\rho)\right]' \left(\tilde{W}_n\right)^{-1} \left[\tilde{m}_n - \widehat{m}_N(\rho)\right]$

- must compute the estimator

$$\widehat{\rho}_n = \underset{\rho}{\text{argmin}}\ s_n(\rho)$$

- an estimate of the Hessian

$$\mathcal{J} = \frac{\partial}{\partial\rho\partial\rho'}\, s^o(\rho)$$

- an estimate of the information

$$\mathcal{I} = \text{Var}\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right] = \mathcal{E}\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right]\left[\frac{\partial}{\partial\rho'}\sqrt{n}\, s_n(\rho^o)\right]'$$

- and an estimate of the variance of $\sqrt{n}(\widehat{\rho}_n - \rho^o)$

$$V_n = \text{Var}\left[\sqrt{n}(\widehat{\rho}_n - \rho^o)\right] = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$$

Computational Strategy – $\hat{\rho}$ & $\hat{\mathcal{J}}^{-1}$

- Chernozhukov, Victor, and Han Hong (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115, 293–346.

- Put $\ell(\rho) = e^{-n\,s_n(\rho)}$. Apply Bayesian MCMC methods with $\ell(\rho)$ as the likelihood and $\pi(\rho,\psi) = \pi_1(\rho)\pi_2(\rho)\pi_3(\rho,\psi)$ as the prior.

- From the resulting MCMC chain $\{\rho_i\}_{i=1}^{R}$, put

$$\hat{\rho}_n = \underset{\rho_i}{\text{argmax}}\, \ell(\rho_i)\pi(\rho_i,\psi^i) \text{ or } \hat{\rho}_n = \bar{\rho}_R = \frac{1}{R}\sum_{t=1}^{R}\rho_i$$

  i.e. the mode or the mean, and put

$$\hat{\mathcal{J}}^{-1} = \left(\frac{n}{R}\right)\sum_{t=1}^{R}\left(\rho_i - \bar{\rho}_R\right)\left(\rho_i - \bar{\rho}_R\right)'$$

Metropolis-Hastings MCMC Chain

Proposal density: $T(\rho_{here}, \rho_{there})$

Proposal: ρ_{prop} drawn from $T(\rho_{old}, \rho)$

Simulate: Get $s_n(\rho_{prop})$, ψ_{prop}, and $\pi(\rho_{prop}, \psi_{prop})$

Likelihood: Put $\ell(\rho) = e^{-n\, s_n(\rho)}$

Put ρ_{new} to ρ_{prop} with probability

$$\alpha = \min\left[1, \frac{\pi(\rho_{prop}, \psi_{prop})\ell(\rho_{prop})T(\rho_{prop}, \rho_{old})}{\pi(\rho_{old}, \psi_{old})\ell(\rho_{old})T(\rho_{old}, \rho_{prop})}\right]$$

Put ρ_{new} to ρ_{old} with probability $1 - \alpha$.

Computational Strategy – $\widehat{\mathcal{I}}$

- For ρ set to $\hat{\rho}_n$, simulate the model and generate I independent data sets $\{\hat{y}_{t,i}\}_{t=1}^{n}$, $i = 1, \ldots, I$, each of exactly the same size n of the original data.

- Let $\hat{s}_{n,i}(\rho)$ denote the criterion function corresponding to data set $\{\hat{y}_{t,i}\}_{t=1}^{n}$. (Store in C++ STL vector indexed by i.)

- Compute $\frac{\partial}{\partial \rho'} \sqrt{n}\, \hat{s}_{n,i}(\bar{\rho}_R)$.

- An estimate of the information is

$$\widehat{\mathcal{I}} = \frac{1}{I} \sum_{i=1}^{I} \left[\frac{\partial}{\partial \rho'} \sqrt{n}\, \hat{s}_{n,i}(\bar{\rho}_R) \right] \left[\frac{\partial}{\partial \rho'} \sqrt{n}\, \hat{s}_{n,i}(\bar{\rho}_R) \right]'$$

EMM Enhancements

Nearly all of the computational cost of the MCMC chain is due to solving the asset pricing equations and computing the criterion function $s_n(\rho)$. This cost can be minimized as follows:

- Reject immediately if $\pi_1(\rho) = 0$.
- Put ρ on a grid. Grid increments determined by sensitivity of $\{\hat{y}_t\}_{t=1}^N$ to ρ elements. E.g. 0.001 for g and δ, and 0.5 for γ.
- Store $s_n(\rho)$, ψ, $\pi_2(\rho)$, $\pi_3(\rho, \psi)$ in a C++ STL associative map indexed by ρ.
- Use table lookup to avoid all recomputation.
- The longer the chain, the faster it runs.

The EMM code does all of this; the case study the first only.

Computational Strategy – EMM MCMC

1. Propose: Draw ρ_{prop} from $T(\rho_{old}, \rho)$.

2. Check support: Check $\pi_1(\rho)$. If $\pi_1(\rho) = 0$, then put ρ_{new} to ρ_{old}. Go to 1.

3. Check map: If ρ_{prop} in map, α can be computed cheaply. Put ρ_{new} to ρ_{prop} with probability α. Put ρ_{new} to ρ_{old} with probability $1 - \alpha$. Go to 1.

4. Simulate: Check $\pi_2(\rho)$. If $\pi_2(\rho) = 0$, then add results to map, put ρ_{new} to ρ_{old}, and go to 1.

5. Evaluate: $s_n(\rho_{prop})$, ψ_{prop}, $\pi(\rho_{prop}, \psi_{prop})$ and put in map. Compute α. Put ρ_{new} to ρ_{prop} with probability α. Put ρ_{new} to ρ_{old} with probability $1 - \alpha$. Go to 1.

Tutorial

- Go through Section 6 of *EMM User's Guide*

- In connection with the files at
  `argux6://home/arg/t/compecon/src/cases/emm/emmrun`