Simulation-Based Estimation with Applications in S-PLUS to Probabilistic Discrete Choice Models and Continuous-Time Financial Models

Eric Zivot

Department of Economics, University of Washington
and Insightful Corporation

Research funded by NSF SBIR Phase II
DMI-0132076

Note: Original PI: Dr. Jiahui Wang, Insightful Corporation

Slides and scripts available at

`http://faculty.washington.edu/ezivot`

1 Efficient Method of Moments

Theory

1. Gallant, A. R., and G. Tauchen (1996). "Which Moments to Match?," *Econometric Theory*

2. Gallant, A. R., and J.R. Long (1997). "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Boimetrika*.

Applications

1. Andersen, T.G. and J. Lund (1997). "Estimating Continuous-Time Stochastic Volatility Models of the Short-term Interest Rate," *Journal of Econometrics*

2. Gallant, A.R., and G. Tauchen (2001). "Efficient Method of Moments," manuscript, Dept. of Economics, University of North Carolina

3. Gallant, A.R., and G. Tauchen (2002). "Simulated Score Methods and Indirect Inference for Continuous-time Models," forthcoming in the *Handbook of Financial Econometric*s.

4. Gallant, R. and G. Tauchen (2001). "SNP: A Program of Nonparametric Time Series Analysis, Version 8.9, User's Guide," available at

`ftp.econ.econ.duke.edu`

5. Gallant, R. and G. Tauchen (2002). "EMM: A Program for Efficient Method of Moments, Version 1.6, User's Guide," available at

`ftp.econ.econ.duke.edu`

1.1 Estimating Dynamic Models with Unobserved States

- Standard statistical methods, both classical and Bayesian, are usually not applicable either because it is not practicable to obtain the density of the state vector or because the integration required to eliminate unobserved states from the likelihood is infeasible.

- On a case-by-case basis, statistical methods are sometimes available, However, the purpose here is to describe methods that are generally applicable.

- Simulating the evolution of the state vector is often practicable. The methods described here rely on this.

- Simulated method of moments methods are described in general, and then the discussion focusses on the efficient method of moments.

Outline of lectures

- Dynamical Systems with Unobserved States
  - Continuous-time diffusion models
  - Example: stochastic volatilty models for interest rates

- Simulated Method of Moments
  - Relationship between GMM and ML
  - Simulated moments estimation using EMM
  - Asymptotics
  - MA(1) example

- The SNP Auxiliary Model

  - Projection
  - Estimation
  - Reprojection

- Simulating from continuous time models
  - Euler's method

- EMM estimation of continuous time models
  - One factor generalized CIR model
  - Two factor model

1.2 General Diffusion Processes

Many continuous time financial models describe the evolution of state variables in terms of a system of stochastic differential equations (SDEs)

$$
\begin{aligned}
U_t &= \text{state variables} \\
dU_t &= A(U_t,\rho)dt + B(U_t,\rho)dW_t \quad 0 \le t < \infty \\
A(U_t,\rho) &= \text{drift} \\
B(U_t,\rho) &= \text{diffusion} \\
\rho &= \text{model parameters} \\
W_t &= \text{Wiener process}
\end{aligned}
$$

The state variables are determined by solving the SDE

$$U_t - U_0 = \int_0^t A(U_s,\rho)ds + \int_0^t B(U_s,\rho)dW_s$$

Observables y_t are regarded as discretely sampled observations from part of the above system. Lo (1988 *ET*) showed that the likelihood of the observed process y_t given ρ is generally not available in closed form.

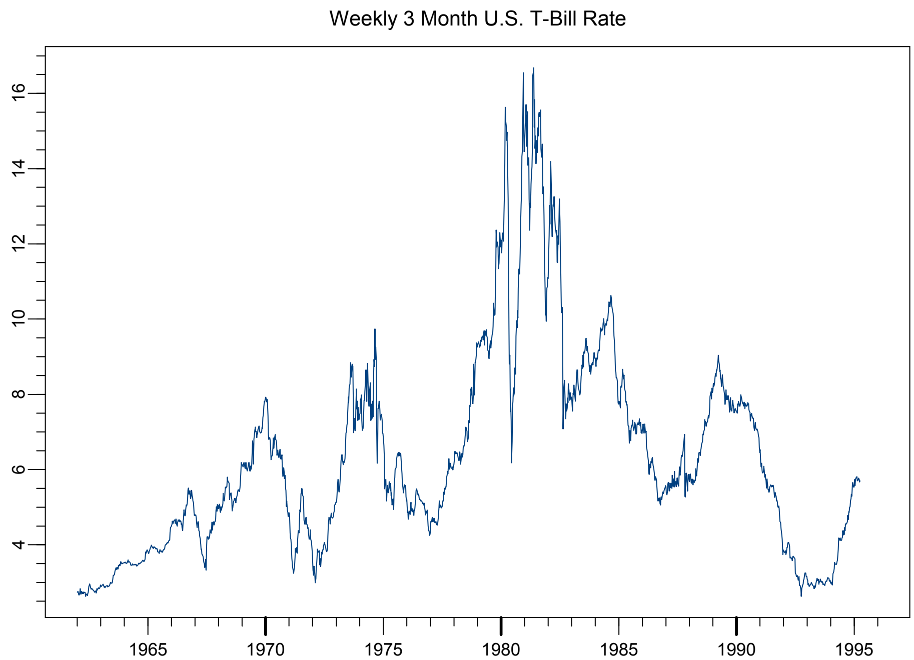

Figure 1:

1.2.1 Example: Short-term Interest Rates

- Weekly (Friday) observations on the 3-month T-Bill rate, 1/5/1962-3/31/1995

- $n = 1,735$ weekly observations

1.2.2 Stochastic Volatility Models of the short-term interest rate

One factor generalized CIR model

$$\begin{aligned}
dr_t &= \kappa(\mu - r_t)dt + \sigma r_t^{\gamma} dW_t \\
U_t &= r_t \\
A(U_t) &= \kappa(\mu - r_t) \\
B(U_t, \rho) &= \sigma r_t^{\gamma} \\
\rho &= (\kappa, \mu, \sigma, \gamma)'
\end{aligned}$$

Andersen, T. and J. Lund (1997). "Estimating Continuous-Time Stochastic Volatility Models of the Short-term Interest Rate," *Journal of Econometrics*.

Two factor model

$$
\begin{aligned}
dv_t &= (a_v + a_{vv}v_t + a_{vr}r_t)dt \\
&\quad +(b_{1v} + b_{1vv}v_t + b_{1vr}r_t)dW_{1t} + b_{2v}dW_{2t} \\
dr_t &= (a_r + a_{rr}r_t)dt + (b_{2r} + b_{2rr}r_t)\exp(v_t)dW_{2t}
\end{aligned}
$$

where

$$
\begin{aligned}
U_t &= (v_t, r_t)', \; W_t = (W_{1t}, W_{2t})' \\
A(U_t, \rho) &= \begin{pmatrix} a_v + a_{vv}v_t + a_{vr}r_t \\ a_r + a_{rr}r_t \end{pmatrix} \\
B(U_t, \rho) &= \begin{pmatrix} b_{1v} + b_{1vv}v_t + b_{1vr}r_t & b_{2v} \\ (b_{2r} + b_{2rr}r_t)\exp(v_t) & 0 \end{pmatrix}
\end{aligned}
$$

Identification restrictions

$$b_{1vv} = 0, b_{2v} = 0, a_v = -a_{vv}$$

1. Gallant, R. and G. Tauchen (1998). "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*.

1.3 ML and GMM Estimation

$$
\begin{aligned}
& y_t \sim iid\ p(y_t|\rho),\ \rho \text{ is } L \times 1 \\
y &= (y_1, \ldots, y_n)' = \text{ random sample}
\end{aligned}
$$

MLE

The MLE for ρ solves

$$\hat{\rho}_{mle} = \arg\min_{\rho} \sum_{t=1}^{n} \ln p(y_t|\rho)$$

Equivalently,

$$
\begin{aligned}
0 &= \sum_{t=1}^{n} \frac{\partial}{\partial \rho} \ln p(y_t|\hat{\rho}_{mle}) \\
&= \sum_{t=1}^{n} s(y_t|\hat{\rho}_{mle})
\end{aligned}
$$

Under regularity conditions, $\hat{\rho}_{mle}$ is consistent and asymptotically efficient

$$\hat{\rho}_{mle} \xrightarrow{p} \rho_0$$
$$\sqrt{n}(\hat{\rho}_{mle} - \rho_0) \xrightarrow{d} N(0, I_0^{-1})$$
$$I_0 = -E\left[\frac{\partial^2 \ln p(y_t|\rho_0)}{\partial\rho\partial\rho'}\right] = E[s(y_t|\rho_0)s(y_t|\rho_0)']$$

GMM

Suppose there are $K \geq L$ moment functions

$$m(y_t, \rho) = (m_1(y_t, \rho), \ldots, m_K(y_t, \rho))'$$

such that

$$E_{\rho_0}[m(y_t, \rho)] = 0$$

For example,

$$\begin{aligned} m_1(y_t, \rho) &= y_t - \mu \\ m_2(y_t, \rho) &= y_t^2 - (\sigma^2 - \mu^2) \end{aligned}$$

The method of moments estimator of ρ is based on matching sample moments to population moments, and solves

$$m_n(\rho) = \frac{1}{n}\sum_{t=1}^{n} m(y_t, \rho) = 0$$

If $K > L$, there is no solution to the above equation. For a symmetric and positive definite weight matrix W, the GMM estimator of ρ solves

$$\hat{\rho}_{GMM} = \arg\min_{\rho}\; m_n(\rho)' W m_n(\rho)$$

Under regularity conditions $\hat{\rho}_{GMM}$ is consistent and asymptotically normally distributed. The efficient GMM estimator uses a weight matrix that is the inverse of a consistent estimate of the asymptotic variance of $\sqrt{n} m_n(\rho_0)$:

$$W_E = \widehat{avar}(\sqrt{n} m_n(\rho_0))^{-1}$$

Remarks:

- The efficient GMM estimator uses the best weight matrix for a given set of moment conditions.

- A different question is: What are the moment conditions such that the GMM estimator is efficient in the class of consistent and asymptotically normal estimators?

Relationship between MLE and GMM

If

$$m_n(\rho) = \frac{1}{n} \sum_{t=1}^{n} s(y_t|\rho)$$

then the GMM estimator is equivalent to MLE. That is, the best moment conditions to use are based on the score of the true model.

Remark

The efficient GMM weight matrix in this case is a consistent estimate of

$$avar(\sqrt{n}m_n(\rho_0))^{-1} = I_0^{-1}$$
$$= \left(E[s(y_t|\rho_0)s(y_t|\rho_0)']\right)^{-1}$$

1.4 Simulated Method of Moments

Notation

$$\begin{aligned} \text{Random Variables} &: \quad y_t,\ t = -\infty, \ldots, \infty \\ \text{Data} &: \quad \tilde{y}_i,\ i = -L, \ldots, n \\ \text{Simulation} &: \quad \hat{y}_i(\rho),\ i = -L, \ldots, N \end{aligned}$$

Dummy arguments of summation

$$(y_{t-L}, \cdots, y_{t-1}, y_t)$$
$$y_t,\ x_{t-1} = (y_{t-L}, \cdots, y_{t-1})$$

Dummy arguments of integration (put $t = 0$)

$$(y_{-L}, \cdots, y_{-1}, y_0)$$
$$y \ = \ y_0,\ x = (y_{-L}, \cdots, y_{-1})$$

Assumption underlying the methodology

For ρ in the parameter space, the model generating the data is presumed to be stationary, ergodic and convenient to simulate

Consequence

For any lag length L and parameter setting ρ there exists a stationary density for observables

$$(y_{t-L}, \dots, y_t) \sim p(y_{-L}, \dots, y_0 | \rho)$$

And, an unconditional expectation

$$E_\rho \left[\psi(y_{-L}, \dots, y_0) \right]$$

can be computed by generating a long simulation

$$\{\hat{y}_t(\rho)\}_{t=-L}^{N}$$

and averaging

$$E_\rho \left[\psi(y_{-L}, \dots, y_0) \right] \approx \frac{1}{N} \sum_{t=0}^{N} \psi(\hat{y}_{t-L}(\rho), \dots, \hat{y}_t(\rho))$$

1.4.1 EMM Estimator

For a quasi-maximum likelihood estimator of an auxiliary model with conditional density $f(\tilde{y}_t|\tilde{x}_{t-1},\theta)$

$$\tilde{\theta}_n = \arg\max_{\theta} \frac{1}{n}\sum_{t=1}^{n} \ln f(\tilde{y}_t|\tilde{x}_{t-1},\theta)$$

a sample average satisfies

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

because these are the first order conditions of the optimization problem

Therefore a large simulation from the model with parametes ρ should satisfy the estimating equations

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

except for sampling variation in $\tilde{\theta}_n$. These estimating equations hold exactly in the limit as n and N tend to infinity.

The EMM estimator attempts to find the parameter vector ρ that solves these estimating equations in a sense to be made precise below.

1.4.2 Minimum Chi-Squared Estimators

If the equations

$$m(\rho, \tilde{\theta}_n) = 0$$

cannot be solved because the dimension of θ is larger than the dimension of ρ, then

- Use a nonlinear optimizer to solve

$$\begin{aligned} \hat{\rho}_n &= \arg\min_{\rho}\ m(\rho, \tilde{\theta}_n)' \left(\tilde{\mathcal{I}}_n\right)^{-1} m(\rho, \tilde{\theta}_n)' \\ \left(\tilde{\mathcal{I}}_n\right)^{-1} &= \text{weight matrix} \end{aligned}$$

- $\mathcal{I}_n$ estimates the asymptotic variance of $\sqrt{n}m(\rho, \tilde{\theta}_n)$. If $f(y|x,\theta)$ is a good approximation to the true data generating process $p(y|x,\rho_0)$ then an adequate estimator is

$$\tilde{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{N}\left[\frac{\partial}{\partial\theta}\ln f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}\ln f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)\right]'$$

Otherwise a HAC estimator of the asymptotic variance should be used.

1.4.3 Asymptotics

Let ρ_0 denote the true value and let θ_0 denote an isolated solution to $m(\rho_0, \theta) = 0$. If $f(y|x, \theta)$ encompasses $p(y|x, \rho_0)$, then under regularity conditions

$$\hat{\rho}_n \xrightarrow{p} \rho \text{ a.s.}$$
$$\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{d} N(0, \left[M_0' \mathcal{I}_0^{-1} M_0\right]^{-1})$$

where

$$\lim_{n\to\infty} \hat{M}_n = M_0$$
$$\lim_{n\to\infty} \tilde{\mathcal{I}}_n = I_0$$
$$\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n), M_0 = M(\rho_0, \theta_0)$$
$$M(\rho, \theta) = \frac{\partial}{\partial \rho'} m(\rho, \theta)$$
$$\mathcal{I}_0 = E_{\rho_0}\left[\left(\frac{\partial}{\partial\theta} \ln f(y_0|x_{-1}, \theta_0)\right)\left(\frac{\partial}{\partial\theta} \ln f(y_0|x_{-1}, \theta_0)\right)'\right]$$

Result: The better the auxiliary model approximates the true structural model, the closer $\hat{\rho}_n$ is to the maximum likelihood estimator based on the true structural model.

1.4.4 $(1-\alpha)\cdot 100\%$ Confidence Intervals

- Wald-type intervals

$$\hat{\rho}_{i,n} \pm z_{\alpha/2} \cdot SE(\hat{\rho}_{i,n})$$

$$SE(\hat{\rho}_{i,n}) \;=\; \sqrt{\left[nM_n'\mathcal{I}_n^{-1}M_n\right]_{ii}^{-1}}$$

Always symmetric - may be misleading

- Invert EMM objective function

$$\begin{aligned} q_i(\rho_i) &= n \cdot \max_{\rho,\rho_i \text{ fixed}} \; m(\rho, \tilde{\theta}_n)' \left(\tilde{\mathcal{I}}_n\right)^{-1} m(\rho, \tilde{\theta}_n)' \\ &= \text{profile objective function} \\ \{\rho_i \;\; &: \;\; q_i(\rho) - q_i(\hat{\rho}_i) \leq \chi^2_{1-\alpha}(1)\} \end{aligned}$$

Better captures nonlinearity of EMM objective function

1.4.5 Specification Test and Diagnostics

Under the null hypothesis that $p(y_{-L}, \ldots, y_0|\rho)$ is the correct model

$$\begin{aligned} L(\hat{\rho}_n, \tilde{\theta}_n) &= n \cdot m(\hat{\rho}_n, \tilde{\theta}_n)' \left(\tilde{\mathcal{I}}_n\right)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n)' \sim \chi^2(q) \\ q &= \dim(\theta) - \dim(\rho) \end{aligned}$$

and

$$\sqrt{n} m(\hat{\rho}_n, \tilde{\theta}_n) \xrightarrow{d} N(0, \mathcal{I}_0 - M_0 \left[M_0' \mathcal{I}_0^{-1} M_0\right]^{-1} M_0')$$

Remarks:

- $L(\hat{\rho}_n, \tilde{\theta}_n)$ is the EMM equivalent of the GMM $J-$statistic. It is an omnibus statistics for misspecification. Large values indicate model misspecification

- Inspection of the $t-$ ratios

$$\begin{aligned} T_n(\hat{\rho}_n, \tilde{\theta}_n) &= \hat{S}_n^{-1}\sqrt{n}m(\hat{\rho}_n, \tilde{\theta}_n) \\ \hat{S}_n &= \left[diag(\tilde{\mathcal{I}}_n - \hat{M}_n \left[\hat{M}_n'\tilde{\mathcal{I}}_n^{-1}\hat{M}_n\right]^{-1} \hat{M}_n')\right]^{1/2} \end{aligned}$$

can suggest reasons for model failure. Different elements of the score correspond to different characteristics of the data. Large $t-$ ratios reveal those characteristics that are not well approximated.

- The $t-$ratios $T_n(\hat{\rho}_n, \tilde{\theta}_n)$ do not account for uncertainty in $\hat{\rho}_n$ and are biased downward (toward zero). They are therefore conservative.

1.4.6 Example: Estimating MA(1) model using EMM

$$\begin{aligned}
y_t &= \varepsilon_t - \gamma\varepsilon_{t-1},\ t = 1, \ldots, n \\
& \quad \varepsilon_t \sim iid\ N(0, \sigma_\varepsilon^2) \\
\rho &= (\gamma, \sigma_\varepsilon)',\ \rho_0 = \text{true value}
\end{aligned}$$

Note: Easy to simulate from MA(1) given ρ.

Auxilary Model

If $|\theta| < 1$ then MA(1) is invertible and has the AR(∞) representation

$$\begin{aligned}
y_t &= \sum_{j=1}^{\infty} \phi_j y_{t-j} + \varepsilon_t \\
\phi_j &= \gamma^j
\end{aligned}$$

A feasible auxilary model is an AR(p) model

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t \\
& \quad u_t \sim iid\ N(0, \sigma_u^2) \\
\theta &= (\phi_1, \ldots, \phi_p, \sigma_u^2)'
\end{aligned}$$

The log-density for the auxilary model is

$$\begin{aligned}\ln f(y_t|x_{t-1},\theta) &= -\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(\sigma_u^2) \\ &\quad -\frac{1}{2\sigma_u^2}(y_t-\phi_1 y_{t-1}-\cdots-\phi_p y_{t-p})^2\end{aligned}$$

If $p \to \infty$ at rate $n^{1/3}$ then $AR(p)$ will encompass the MA(1) data generating process as $n \to \infty$.

Quasi-maximum likelihood estimation

Given a sample of observed data $\{\tilde{y}_t\}_{t=1}^{n}$ based on ρ_0, quasi-maximum likelihood estimation of θ solves

$$\tilde{\theta}_n = \arg\max_{\theta} \frac{1}{n}\sum_{t=1}^{n} \ln f(\tilde{y}_t|\tilde{x}_{t-1},\theta)$$

The sample score vector

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

has elements

$$\frac{1}{n}\sum_{t=1}^{n}(\tilde{y}_t - \tilde{\phi}_1\tilde{y}_{t-1} - \cdots - \tilde{\phi}_p\tilde{y}_{t-p})\tilde{y}_{t-1}$$

$$\vdots$$

$$\frac{1}{n}\sum_{t=1}^{n}(\tilde{y}_t - \tilde{\phi}_1\tilde{y}_{t-1} - \cdots - \tilde{\phi}_p\tilde{y}_{t-p})\tilde{y}_{t-p}$$

$$-\frac{1}{2\tilde{\sigma}^2}[n + \tilde{\sigma}^{-2}]\sum_{t=1}^{n}(\tilde{y}_t - \tilde{\phi}_1\tilde{y}_{t-1} - \cdots - \tilde{\phi}_p\tilde{y}_{t-p})$$

and by construction $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n) = 0$.

For a fixed ρ, let $\{\hat{y}_t(\rho)\}_{t=1}^{N}$ denote a simulated sample of size $N >> n$ from the MA(1) model. The sample score of the auxiliary model evaluated using the simulated data is

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

which has elements

$$\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t(\rho)-\tilde{\phi}_1\hat{y}_{t-1}(\rho)-\cdots-\tilde{\phi}_p\hat{y}_{t-p}(\rho))\hat{y}_{t-1}(\rho)$$

$$\vdots$$

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t(\rho)-\tilde{\phi}_1\hat{y}_{t-1}(\rho)-\cdots-\tilde{\phi}_p\hat{y}_{t-p}(\rho))\hat{y}_{t-p}(\rho)$$

$$-\frac{1}{2\hat{\sigma}^2}[n+\tilde{\sigma}^{-2}]\sum_{t=1}^{n}(\hat{y}_t(\rho)-\tilde{\phi}_1\hat{y}_{t-1}(\rho)-\cdots-\tilde{\phi}_p\hat{y}_{t-p}(\rho))$$

If $\rho=\rho_0$ then

$$m(\rho_0,\tilde{\theta}_n)\approx 0$$

whereas if $\rho\neq\rho_0$ then

$$m(\rho,\tilde{\theta}_n)\neq 0$$

Since θ is $(p+1)\times 1$ and ρ is (2×1) it is not possible to find $\hat{\rho}$ such that $m(\hat{\rho},\tilde{\theta}_n)=0$. Instead ρ is estimated by minimizing the quadratic form

$$\hat{\rho}_n=\arg\min_{\rho}\ m(\rho,\tilde{\theta}_n)'\left(\tilde{\mathcal{I}}_n^{-1}\right)m(\rho,\tilde{\theta}_n)$$

1.4.7 EMM Estimation using SPLUS.

Example Data

$$\text{Structural model: } MA(1)$$
$$y_t = \varepsilon_t + 0.75\varepsilon_{t-1},\ \varepsilon_t \sim iid\ N(0,1)$$
$$\rho_0 = (-0.75, 1)'$$
$$n = 250$$
$$\text{Auxilary model: } AR(3)$$
$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + u_t,$$
$$u_t \sim iid\ N(0, \sigma_u^2)$$
$$\theta = (\mu, \phi_1, \phi_2, \phi_3, \sigma_u^2)'$$
$$N = 10,000$$
$$\dim(\theta) - \dim(\rho) = 3$$

1.5 General Purpose Auxiliary Model: SNP

- SNP is a Semi-NonParametric method, based on an expansion in Hermite functions, for estimation of the one-step-ahead conditional density $f(y_t|x_{t-1}, \theta)$.

- It consists of applying classical *parametric* estimation and inference procedures to models derived from *nonparametric* truncated series expansions.

- Estimation of SNP models entails using a standard maximum likelihood procedure together with a model selection strategy that determines the appropriate degree of the nonparametric expansion. Under reasonable regularity conditions, the estimator is consistent and efficient.

- An important usage of SNP models is to act as a score generator for EMM estimation.

1.5.1 Assumptions

1. Stationary multivariate data

$$y_t = (y_{1t}, \ldots, y_{Mt})'$$

   Stationarity is assumed so that densities for a stretch of data are time invariant.

2. Markovian structure: the conditional density of y_t given the entire past depend only on a finite number of lags $x_{t-1} = (y_{t-L}, \ldots, y_{t-1})$

3. Smoothness: the density $f(y_t, x_{t-1})$ must have derivatives to the order $ML/2$ or higher and have tails that are bounded above by

$$P(y_{t-L}, \ldots, y_t) \exp\left(\frac{1}{2}\sum_{t=0}^{L} y_{t-L}^2\right)$$

   where P is a polynomial of large but finite degree.

4. Focus is on the estimation of the transition density

$$f(y_t|x_{t-1}) = f(y_t|y_{t-L}, \dots, y_{t-1})$$

1.5.2 Overview of SNP Model

Location-scale transformation for a multivariate innovation z

$$y = R_x z + \mu_x$$
$$R \text{ is upper triangular}$$
$$z = R_x^{-1}(y - \mu_x)$$

Innovation density (Hermite expansion)

$$\begin{aligned} h_K(z|x) &= \frac{[P(z,x)]^2\,\phi(z)}{\int [P(u,x)]^2\,\phi(u)du} \\ \phi(z) &= N(0, I_M) \end{aligned}$$

$P(z,x) =$ polynomial in z of degree K whose coefficients are polynomials of degree K_x in x

Location function (VAR)

$$\begin{aligned} \mu_x &= b_0 + Bx_{t-1} \\ x_{t-1} &= (y_{t-L}, \ldots, y_{t-1}) \end{aligned}$$

ARCH scale function

$$vech(R_{x_{t-1}}) = r_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}|$$

Garch scale function

$$\begin{aligned} vech(R_{x_{t-1}}) &= r_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}| \\ &\quad + \sum_{i=1}^{L_g} diag(G_{(i)}) R_{x_{t-2-L_g+i}} \end{aligned}$$

DVEC(1,1) Garch option

1.5.3 Notation for Multivariate Polynomial

Degree K, dimension M

$$\begin{aligned} P(z) &= \sum_{|\alpha|=0}^{K} a_\alpha z^\alpha \\ z^\alpha &= z_1^{\alpha_1} \cdot z_2^{\alpha_2} \cdot \dots \cdot z_M^{\alpha_M} \\ |\alpha| &= |\alpha_1| + |\alpha_2| + \dots + |\alpha_M| \end{aligned}$$

Example: $K = 2, M = 2$

$$\begin{aligned} z &= (z_1, z_2)', \ \alpha = (\alpha_1, \alpha_2)' \\ |\alpha| &= 0 \Rightarrow \alpha = (0, 0) \\ |\alpha| &= 1 \Rightarrow \alpha = (1, 0)', (0, 1)' \\ |\alpha| &= 2 \Rightarrow \alpha = (1, 1)', (2, 0)', (0, 2)' \end{aligned}$$

and

$$\begin{aligned} P(z) &= a_{(0,0)} + a_{(1,0)} z_1 + a_{(0,1)} z_2 \\ &\quad + a_{(1,1)} z_1 z_2 + a_{(2,0)} z_1^2 + a_{(0,2)} z_2^2 \end{aligned}$$

1.5.4 Hermite Expansions: Rationale 1

An unnormalized Hermite polynomial has the form

$$P(z)\sqrt{\phi(z)}$$

$$\phi(z) = N(0, I_M) = (2\pi)^{-M/2} \exp\left(-\frac{1}{2}z'z\right)$$

A function $g(z)$ that satisfies

$$||g||_2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g^2(z)dz_1 \cdots dz_M < \infty$$

is called an L_2 function.

Result: The Hermite polynomials are dense in $L_2(-\infty, \infty)$; i.e.,

$$\lim_{K\to\infty} ||g(z) - P(z)\sqrt{\phi(z)}||_2 = 0$$

where the coefficients $\{a_\alpha\}_{|\alpha|<K}$ of $P(z)$ are those that minimize the approximation error $||g(z) - P(z)\sqrt{\phi(z)}||_2$.

1.5.5 Hermite Expansions: Rationale 2

Let $h(z)$ be a density function. Because

$$\int h(z)dz = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(z)dz_1 \cdots dz_M = 1$$

$\sqrt{h(z)}$ is in $L_2(-\infty, \infty)$ and can therefore be approximated by $P(z)\sqrt{\phi(z)}$ as accurately as desired by taking K large enough.

This fact motivates using

$$h_K(z) = \frac{P^2(z)\phi(z)}{\int P^2(s)\phi(s)ds}$$

to approximate $h(z)$, where the division is to guarantee that $h_K(z)$ integrates to 1.

1.5.6 The Main Idea

Take $h(z)$ as the parent density and use a location-scale transform

$$y = Rz + \mu$$

to generate a location-scale family of densities

$$f(y|\theta) = \frac{\left\{P[R^{-1}(y-\mu)]\right\}^2 \phi[R^{-1}(y-\mu)]}{|\det(R)| \int P^2(s)\phi(s)ds}$$

which can be estimated from data $\{\tilde{y}_t\}_{t=1}^n$ by quasi maximum likelihood

$$\tilde{\theta} = \arg\max_{\theta} \prod_{t=1}^{n} f(\tilde{y}_t|\theta)$$

The density estimate is

$$f(y) = f(y|\tilde{\theta})$$

The consistency of the estimator was established by Gallant and Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*.

1.5.7 SNP Density: iid Data

Location-scale transform

$$
\begin{aligned}
y &= Rz + \mu,\ z = R^{-1}(y - \mu) \\
R &= \text{upper triangular}
\end{aligned}
$$

Density

$$
\begin{aligned}
f(y|\theta) &\propto P^2[R^{-1}(y - \mu)] \cdot N(y|\mu, \Sigma) \\
\Sigma &= RR'
\end{aligned}
$$

Example: $K = 2$, $M = 2$

$$
\begin{aligned}
R &= \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \\
\theta &= (a_{(0,0)}, a_{(1,0)}, a_{(0,1)}, \\
&\quad a_{(1,1)}, a_{(2,0)}, a_{(0,2)}, \\
&\quad \mu_1, \mu_2, R_{11}, R_{12}, R_{22}) \\
a_{(0,0)} &= 1 \text{ (normalization)}
\end{aligned}
$$

How well does SNP do at approximating densities?

- SNP performs well relative to kernel estimators
- SNP performs well on the Marron and Wand (1992) test suite densities.

References

1. Fenton, V.M. and R. Gallant (1996). Convergence Rates of SNP Density Estimators, *Econometrica*.

2. Fenton, V.M. and R. Gallant (1996). Qualitative and Asymptotic Performance of SNP Density Estimators, *Journal of Econometrics*.

1.5.8 Choice of K

Results in Coppejans and Gallant (2000), "Cross Validated SNP Density Estimates," indicate that minimizing the BIC seems to work well for determining the degree K of the polynomial $P(z)$

Estimation: Eqivalent to maximum likelihood, but more numerically stable is to minimize the negative of the average log likelihood

$$\begin{aligned} \tilde{\theta}_n &= \operatorname*{arg\,min}_{\theta} \; s_n(\theta) \\ s_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\theta) \end{aligned}$$

Schwarz criterion (BIC): Choose K to minimize

$$\begin{aligned} BIC(p) &= s_n(\theta) + \frac{p}{2n}\ln(n) \\ p &= \text{number of free parameters in } \theta \end{aligned}$$

1.5.9 SNP Transition Density for Time Series

The idea is to modify the location and scale transforms of the SNP density for iid data to be functions of the past, which is the standard method of modifying a model for iid data for application to time series data. Lastly, the SNP density itself is modified to accomodate non-homogeneous innovations. The steps are as follows.

VAR location function:

$$\begin{aligned} y &= Rz + \mu_{x_{t-1}} \\ \mu_{x_{t-1}} &= b_0 + Bx_{t-1} \\ x_{t-1} &= (y_{t-L_u}, \ldots, y_{t-1})' \\ L_u &= \text{VAR order} \end{aligned}$$

b_0 is $M \times 1$, B is $M \times L_u$.

Density

$$\begin{aligned} f(y|\theta) &\propto P^2[R^{-1}(y - \mu_{x_{t-1}})] \cdot N(y|\mu, RR') \\ K_z &= 0 \Rightarrow \text{Gaussian VAR} \\ K_z &> 0 \Rightarrow \text{non-Gaussian VAR} \end{aligned}$$

ARCH-type scale function

$$\begin{aligned}
y &= R_{x_{t-1}} z + \mu_{x_{t-1}} \\
\mu_{x_{t-1}} &= b_0 + B x_{t-1} \\
vech(R_{x_{t-1}}) &= r_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}| \\
L_r &= \text{ARCH order}
\end{aligned}$$

where

$$\begin{aligned}
& r_0 \text{ is } M(M+1)/2 \times 1 \\
P &= [P_{(1)}|\cdots|P_{(L_r)}] \text{ is } M(M+1)/2 \times L_r
\end{aligned}$$

Density

$$\begin{aligned}
f(y|\theta) &\propto P^2[R_{x_{t-1}}^{-1}(y - \mu_{x_{t-1}})] \cdot N(y|\mu, R_{x_{t-1}} R'_{x_{t-1}}) \\
K_z &= 0, L_r > 0 \Rightarrow \text{Gaussian VAR-ARCH} \\
K_z &> 0, L_r > 0 \Rightarrow \text{non-Gaussian VAR-ARCH}
\end{aligned}$$

GARCH-type scale function

$$
\begin{aligned}
y &= R_{x_{t-1}} z + \mu_{x_{t-1}} \\
\mu_{x_{t-1}} &= b_0 + B x_{t-1} \\
vech(R_{x_{t-1}}) &= r_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}| \\
&\quad + \sum_{i=1}^{L_g} diag(G_{(i)}) R_{x_{t-2-L_g+i}} \\
L_g &= \text{ARCH order}
\end{aligned}
$$

where

$$
G = [G_{(1)}|\cdots|G_{(L_g)}] \text{ is } M(M+1)/2 \times L_g
$$

Density

$$
\begin{aligned}
f(y|\theta) &\propto P^2[R_{x_{t-1}}^{-1}(y - \mu_{x_{t-1}})] \cdot N(y|\mu, R_{x_{t-1}} R'_{x_{t-1}}) \\
K_z &= 0, L_g > 0 \Rightarrow \text{Gaussian VAR-GARCH} \\
K_z &> 0, L_r > 0 \Rightarrow \text{non-Gaussian VAR-GARCH}
\end{aligned}
$$

1.5.10 SNP for Non-homogenous Innovations

Non-homogeneous innovations are accomodated by letting the polynomial part of the SNP model

$$P(z) = \sum_{|\alpha|=0}^{K_z} a_\alpha z^\alpha$$

have coefficients a_α that are polynomials in the conditioning information

$$x = (x_{t-L_p}, \ldots, x_{t-1})$$

giving

$$a_\alpha(x) = \sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta$$

The polynomial part is denoted

$$P(z,x) = \sum_{|\alpha|=0}^{K_z} \sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta z^\alpha$$

The SNP density for the non-homogeneous innovations is a Hermite polynomial in z whose coefficients are polynomials in x

$$
\begin{aligned}
h_K(z|x) &= \frac{P(z,x)\phi(z)}{\int P(s,x)\phi(s)ds} \\
K &= (K_z, K_x)'
\end{aligned}
$$

1.5.11 SNP: Putting It All Together

The general non-Gaussian VAR model with GARCH errors is

$$
\begin{aligned}
y &= R_{x_{t-1}} z + \mu_{x_{t-1}} \\
\mu_{x_{t-1}} &= b_0 + B x_{t-1} \\
vech(R_{x_{t-1}}) &= r_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}| \\
&\quad + \sum_{i=1}^{L_g} diag(G_{(i)}) R_{x_{t-2-L_g+i}}
\end{aligned}
$$

with density

$$
\begin{aligned}
f(y|\theta) &\propto P^2[R^{-1}_{x_{t-1}}(y - \mu_{x_{t-1}}), x_{t-1}] \cdot N(y|\mu, R_{x_{t-1}} R'_{x_{t-1}}) \\
K_z &> 0 \Rightarrow \text{non-Gaussian VAR-GARCH} \\
K_z &> 0, K_x > 0 \Rightarrow \text{non-Gaussian,} \\
&\quad \text{non-linear VAR-GARCH} \\
L_p &= \text{number of lags of } x_{t-1} \text{ in } P(z, x_{t-1})
\end{aligned}
$$

1.5.12 Data Transformations

Standardizing data

$$
\begin{aligned}
\bar{y} &= \frac{1}{n}\sum_{t=1}^{n} \tilde{y}_t \\
S &= \frac{1}{n}\sum_{t=1}^{n} (\tilde{y}_t - \bar{y})(\tilde{y}_t - \bar{y})' \\
\tilde{y}_t^* &= diag(S^{-1/2})(\tilde{y}_t - \bar{y}) \\
S^{-1/2} &= \text{Choleski factor}
\end{aligned}
$$

Gallant and Tauchen suggest using only diagonal elements of S

Adjusting for extreme values in x_{t-1}

Problem: If the ture density $f(y|x)$ is heavy tailed, then x_{t-1} will contain extreme observations, which have a strong and undesirable influence on estimates when $L_r > 0$ (ARCH terms). Additionally, if data are highly persistent (like interest rates), fitted SNP models can generate explosive simulations

Cure: Replace each component of x by its log spline transform

$$\begin{aligned}
\tilde{x}_i^* &= \begin{cases} \frac{1}{2}\left[\tilde{x}_i - \sigma_{tr} - \ln(1 - \tilde{x}_i - \sigma_{tr})\right] & \tilde{x}_i < -\sigma_{tr} \\ x_t & -\sigma_{tr} \leq \tilde{x}_i \leq \sigma_{tr} \\ \frac{1}{2}\left[\tilde{x}_i + \sigma_{tr} + \ln(1 - \sigma_{tr} + \tilde{x}_i)\right] & \sigma_{tr} < \tilde{x}_i \end{cases} \\
\sigma_{tr} &= 4
\end{aligned}$$

Note: Spline transformation attenuates potential explosive simulations $\hat{y}_t(\rho)$ that may occur during course of EMM estimation

Order in Which Transformations are Applied

$$\tilde{y}_t \text{ (raw data)}$$
$$\downarrow$$
$$\tilde{y}_t^* \text{ (centered and scaled)}$$
$$\downarrow$$
$$\tilde{x}_{t-1} \text{ (lags)}$$
$$\downarrow$$
$$\tilde{x}_{t-1}^* \text{ (spline)}$$
$$\downarrow$$
$$\mu_x,\ R_x \text{ (location-scale)}$$

1.5.13 Asymptotics

If the parameters of $f(y|x,\theta)$ are estimated by quasi-maximum likelihood

$$\begin{aligned} \tilde{\theta}_n &= \operatorname*{arg\,min}_{\theta} s_n(\theta) \\ s_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n} \ln f(\tilde{y}_t|\tilde{x}_{t-1},\theta) \end{aligned}$$

and $K = (K_z, K_x)'$ grows with the sample size, then the estimator

$$f_n(y|x) = f(y|x,\theta)$$

converges almost surely to the true transition density $f(y|x)$ in Sobolev norm as $n \to \infty$. Moreover, K can depend on the data.

Reference:

Gallant, R. and D. Nychka (1987). "Seminonparametric Maximum Likelihood Estimation," Econometrica.

1.5.14 Model Selection

Schwarz criterion (BIC): Choose K to minimize

$$\begin{aligned} BIC(K) &= s_n(\theta) + \frac{p}{2n}\ln(n) \\ s_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\theta) \\ p &= \text{number of free parameters in } \theta \end{aligned}$$

Suggested Search Order

1. Determine best VAR order L_u
2. Determine best ARCH and GARCH orders L_r, L_g
3. Determine best z ploynomial order K_z (start at $K_z = 4$)
4. Determine the best x polynomial order K_x

1.6 Simulation Methods for SDEs

SDE

$$dU_t = A(U_t)dt + B(U_t)dW_t$$

Euler's method: Iterate for small delta (Δ)

$$U_\Delta - U_0 = A(U_0)\Delta + B(U_0)(W_\Delta - W_0)$$

Sum

$$\sum_{i=1}^{t/\Delta} \left[U_{i\Delta} - U_{(i-1)\Delta}\right] = \sum_{i=1}^{t/\Delta} A\left[U_{(i-1)\Delta}\right]\Delta + \sum_{i=1}^{t/\Delta} B\left[U_{(i-1)\Delta}\right]\Delta\left[W_{i\Delta} - W_{(i-1)\Delta}\right]$$

Limit as $\Delta \to 0$

$$U_t - U_0 = \int_0^t A(U_s)ds + \int_0^t B(U_s)dW_s$$

1.6.1 Example: Simulating from interest rate models

CIR Model

$$
\begin{aligned}
dr_t &= \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t \\
A(r_t) &= \kappa(\mu - r_t) \\
B(r_t, \rho) &= \sigma\sqrt{r_t} \\
\rho &= (\kappa, \mu, \sigma)' = (0.1, 0.08, 0.06)' \\
N &= 250 \\
\Delta &= 1/100
\end{aligned}
$$

Two factor interest rate diffusion

$$
\begin{aligned}
dv_t &= (a_v + a_{vv} v_t + a_{vr} r_t)dt \\
&\quad +(b_{1v} + b_{1vv} v_t + b_{1vr} r_t) dW_{1t} + b_{2v} dW_{2t} \\
dr_t &= (a_r + a_{rr} r_t)dt + (b_{2r} + b_{2rr} r_t) \exp(v_t) dW_{2t} \\
b_{1vv} &= 0, b_{2v} = 0, a_v = -a_{vv} \\
\boldsymbol{\rho} &= (a_v, a_{vr}, a_r, a_{rr}, \\
&\quad b_{1v}, b_{1vv}, b_{1vr}, b_{2v}, b_{2r}, b_{2rr}) \\
&= (-0.18, -0.0088, 0.19, -0.0035, \\
&\quad 0.69, 0, -0.063, 0, 0.038, -0.17)'
\end{aligned}
$$

1.7 EMM Estimation of Continuous Time Models for Interest Rates

1. Fit SNP Model to observed data

2. Create simulator functon for continuous time model

3. Estimate model by EMM

4. Check EMM diagnostics