Stochastic Volatility - Likelihood Inference and Comparison with ARCH Models

Jian Zhou, Chi Zhou, Yuting Fang

Peking University

zhou_jian@pku.edu.cn

January 14, 2021

Roadmap

1. Introduction
   - Specification of SV model
   - Motivation
2. Gibbs sampling
   - Sampling details
   - Illustration
3. Improved MCMC algorithms
   - Offset mixture representation
   - Mixture Simulator
   - Integrating out the log-volatilities
   - Re-weighting
4. Model Diagnosis and Comparison
5. SV model extensions
6. Conclusion

Specification of SV model

The variance of returns on assets tends to change over time.

Two ways of modelling changing variance:

- Auto-regressive conditional heteroscedasticity (ARCH)
- Stochastic Volatility (SV)

Canonical SV model

$$
\begin{aligned}
y_t &= \beta e^{h_t/2}\varepsilon_t, t \geq 1, \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \\
h_1 &\sim \mathbb{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)
\end{aligned}
\tag{1}
$$

Motivaton

No MCMC simulation based analysis of the SV model.

- The Kalman filter method (maximize quasi-likelihood) behaves disappointing with small sample size because it uses normal distribution to approximate $log\epsilon^2$ (Harve, Ruiz, and Shephard, 1994)
- Solution: Gibbs sampling algorithm for the SV model

No tools for model diagnosis and comparison

- Solution: Develop useful diagnostics for detecting model failure, and formal tools for comparing models

An initial Gibbs sampling algo for SV model

- This method is suggested by Jacquier, Polson and Rossi (1994) and Shephard (1993)
- The algorithm is defined by a blocking scheme $\psi = (\psi_1, ..., \psi_d)$ and the full conditional distributions $\psi_i|\psi_{\backslash i}$.
- For SV model, the ψ vector becomes (h, θ)

Gibbs sampling

1. Initialize h and θ.
2. Sample h_t from $h_t|h_{\backslash t}, y, \theta, t = 1, ..., n$.
3. Sample $\sigma_\eta^2|y, h, \phi, \mu, \beta$
4. Sample $\phi|h, \mu, \beta, \sigma_\eta^2$
5. Sample $\mu|h, \phi, \sigma_\eta^2$
6. Go to 2.

Prior of h_t

We can sample the posterior $f(h_t|h_{\setminus t}, \theta, y)$ using a simple A-R procedure.

- The update is $f(h_t|h_{\setminus t}, \theta, y) \propto f(h_t|h_{\setminus t}, \theta) f(y_t|h_t, \theta)$

$$
\begin{aligned}
f(h_t|h_{\setminus t}, \theta) &= f(h_t|h_{t-1}, h_{t+1}, \theta) = f_N(h_t|h_t^*, \nu^2) \\
\text{where } h_t^* &= \mu + \frac{\phi[(h_{t-1} - \mu) + (h_{t+1} - \mu)]}{1 + \phi^2} \text{ and } \nu^2 = \frac{\sigma_\eta^2}{1 + \phi^2}
\end{aligned}
\tag{2}
$$

- h_t^* is simply an equivalent expression of $\mathbb{E} h_t$

Likelihood $f(y_t|h_t,\theta)$ and valid $g(\cdot)$ for A-R

- From $y_t = e^{h_t/2}\varepsilon_t$ and $\varepsilon \sim \mathbb{N}(0,1)$, the conditional distribution is $y_t|h_t,\theta \sim \mathbb{N}(0, e^{h_t})$.
- Take log w.r.t. $f(y_t|h_t,\theta)$, $\log f(y_t|h_t,\theta) = \text{const} + \log f^*(y_t,h_t,\theta)$

$$\begin{aligned} \log f^*(y_t,h_t,\theta) &= -\frac{1}{2}h_t - \frac{y_t^2}{2}e^{-h_t} \\ &\leq -\frac{1}{2}h_t - \frac{y_t^2}{2}\left[e^{-h_t^*}(1+h_t^*) - h_t e^{-h_t^*}\right] \\ &= \log g^*(y_t,h_t,\theta,h_t^*) \end{aligned} \tag{3}$$

- Hence, we can sample the LHS by A-R and RHS is normally dist.

$$\begin{gathered} f(h_t|h_{\setminus t},\theta)f^*(y_t,h_t,\theta) \leq f_N(h_t|h_t^*,\nu^2)g^*(y_t,h_t,\theta,h_t^*) \\ RHS \sim \mathbb{N}\left(h_t^* + \frac{\nu^2}{2}\left[y_t^2 e^{-h_t^*} - 1\right], \nu^2\right) \end{gathered} \tag{4}$$

Posteriors – $\sigma_\eta^2 | y, h, \phi, \mu$

Sampling $\sigma_\eta^2 | y, h, \phi, \mu$

We assume a conjugate prior $\sigma_\eta^2 | \phi, \mu \sim \mathcal{IG}(\sigma_r/2, S_\sigma/2)$, by Normal-IG update, σ_η^2 is sampled from:

$$\sigma_\eta^2 | y, h, \phi, \mu \sim \mathcal{IG}\left(\frac{n+\sigma_r}{2}, \frac{S_\sigma + (h_1-\mu)^2(1-\phi^2) + Q}{2}\right) \tag{5}$$

where $\quad Q = \sum_{t=1}^{n-1} \left((h_{t+1}-\mu) - \phi(h_t-\mu)\right)^2$

- In our replication, σ_r is set to be 5 and S_σ is 0.05.

Posteriors – $\phi | h, \sigma_\eta^2, \mu$

Sampling $\phi | h, \sigma_\eta^2, \mu$

Let $\phi = 2\phi^* - 1$ where ϕ^* is distributed as $Beta(\phi^{(1)}, \phi^{(2)})$, the prior is

$$\pi(\phi) \propto \left[\frac{1+\phi}{2}\right]^{\phi^{(1)}-1} \left[\frac{1-\phi}{2}\right]^{\phi^{(1)}-1}, \quad \phi^{(1)}, \phi^{(1)} > \frac{1}{2} \tag{6}$$

this limits ϕ to $(-1, 1)$ so that the AR(1) process of log-vol is stationary. The full conditional density is proportional to $\pi(\phi) f(h|\mu, \phi, \sigma_\eta^2)$, where

$$\log f(h|\cdot) \propto -\frac{(h_1 - \mu)^2(1-\phi^2) + Q}{2\sigma_\eta^2} + \frac{1}{2}\log(1-\phi^2) \tag{7}$$

- The full conditional density is hard to sample, here A-R can be applied.

Posteriors – $\phi | h, \sigma_\eta^2, \mu$

- By taking Taylor expansion at

$$\hat{\phi} = \sum_{t=1}^{n-1}(h_{t+1} - \mu)(h_t - \mu) / \sum_{t=1}^{n-1}(h_t - \mu)^2 \tag{8}$$

- Sample a proposal ϕ^* from $\mathbb{N}(\hat{\phi}, V_\phi)$, where $V_\phi = \sigma_\eta^2 \left[\sum (h_t - \mu)^2\right]^{-1}$
- Accept this proposal value as $\phi^{(i)}$ with probability $\exp\{g(\phi^*) - g(\phi^{(i-1)})\}$, where

$$g(\phi) = \log(\phi) - \frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2}\log(1 - \phi^2) \tag{9}$$

Posteriors – $\mu | h, \phi, \sigma_\eta^2$

- In this work, β is set to 1, and let μ be unrestricted.
- It takes a informative prior: $\mu \sim \mathbb{N}(0, 10)$

Sampling $\mu | h, \phi, \sigma_\eta^2$

The full conditional distribution is Normal, intuitively from Normal-Normal update.

$$\mu | h, \phi, \sigma_\eta^2 \sim \mathbb{N}(\hat{\mu}, \sigma_\mu^2) \tag{10}$$

where

$$\hat{\mu} = \sigma_\mu^2 \left[\frac{1-\phi^2}{\sigma_\eta^2} h_1 + \frac{1-\phi}{\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \right] \tag{11}$$

and

$$\sigma_\mu^2 = \sigma_\eta^2 \left[(n-1)(1-\phi)^2 + 1 - \phi^2 \right]^{-1} \tag{12}$$

Illustration, using GBP/USD

- Initialization: $h_t = 0$, $\phi = 0.95$, $\sigma_\eta^2 = 0.02$and $\mu = 0$.
- The simulation does 1 million sweeps, here are the results

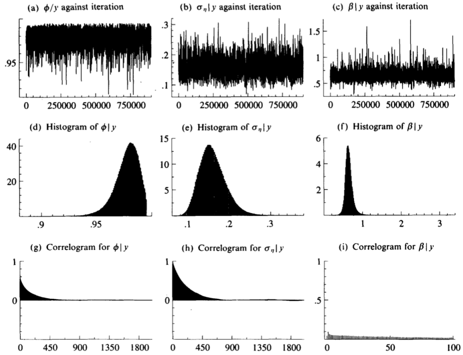

FIGURE 2

Single move Gibbs sampler for the Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 1,000,000 iterations were drawn, discarding the first 50,000

Illustration, using GBP/USD

TABLE 1.

Daily returns for Sterling: summaries of Figure 2. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 2000, 4000 and 2000 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 sweeps of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97762	0·00013754	163·55	0·00011062	*−0·684*	*0·203*
$\sigma_\eta \mid y$	0·15820	0·00063273	386·80	−0·00022570	0·00098303	*−0·129*
$\beta \mid y$	0·64884	0·00036464	12·764	0·00021196	−0·00040183	0·0098569
Time	3829·5	0·58295				

- Simulation Inefficiency is defined by the variance of the sample mean from the MCMC sampling scheme divided by the variance divided by the number of iterations.
- Large Inefficiency reflects poorly mixed chain.
- Long lasting auto-correlation and slow convergence.

Offset mixture representation

Improvement on the previous Gibbs sampler

In this section, an offset mixture of normal distributions is offered to accurately approximate the exact likelihood. The efficiency and precision is greatly improved.

- The parametric model is $\log(y_t^2) = h_t + \log(\varepsilon_t^2)$.
- The offset mixture model is a linear approximation:

$$y_t^* = h_t + z_t \tag{13}$$

  where $y_t^* = \log(y_t^2 + c)$ and $f(z_t) = \sum_i q_i f_N(z_t | m_i - 1.2704, \nu_i^2)$
- The state indicator variable s_t used in MCMC is:

$$z_t | s_t = i \ \sim \mathbb{N}(m_i - 1.2704, \nu_i^2), \ \Pr(s_i = i) = q_i \tag{14}$$

Offset mixture representation

We need to select K and $\{m_i, q_i, \nu_i^2\}$. Since we use z_t to approximate $\log(\varepsilon_t^2)$, the optimization is on a non-linear least square program to match the first fourth moments of $f(z_t)$ and $\log \chi_1^2$, with restrictions that the approximating densities lie with in a small distance of the true density.

The satisfactory answers are $K = 7$ and other parameters in TABLE 4.

TABLE 4

Selection of the mixing distribution to be $\log \chi_1^2$

ω	$\Pr(\omega = i)$	m_i	σ_i^2
1	0·00730	−10·12999	5·79596
2	0·10556	−3·97281	2·61369
3	0·00002	−8·56686	5·17950
4	0·04395	2·77786	0·16735
5	0·34001	0·61942	0·64009
6	0·24566	1·79518	0·34023
7	0·25750	−1·08819	1·26261

Mixture Simulator

Sampler of the mixture distribution

1. Initialize s, ϕ, σ_η^2 and μ.
2. Sample h from $h|y^*, s, \phi, \sigma_\eta^2, \mu$.
3. Sample s from $s|y^*, h$.
4. Update ϕ, σ_η^2, μ according to (7), (5) and (10)
5. Go to 2.

- We can directly sample $h|y^*, s, \phi, \sigma_\eta^2, \mu$, because it is multivariate normal. A-R sampling for h_t in the former algo is dropped.
- Thus, the precision is improved.

Signal smoother draws from filtered normal (Optional.)

- How does Kalman filter work?
  - Knowing all the data $\{y_t\}_{t=1}^{n}$, and the state space model.
  - The filtered density is:

$$p(h_t|y_{1:t}) \propto p(y_t|h_t)p(h_t|y_{1:t-1}) \tag{15}$$

  - The prediction $p(h_t|y_{1:t-1})$ can be further expressed by filtered density at $t-1$.
  - So filtering is an iterative process beginning from the measurement probability, and the results are the best smoothing joint normal posterior density for h_t sampling.
- From the measurement probability $p(y_t|h_t)$, they applied Kalman filter to derive the multivariate normal posterior of h (See Appendix 1):

$$(c_1 + Z_1\gamma_1, ..., c_n + Z_n\gamma_n)|y, \theta \tag{16}$$

- Take draws from (16), we have all h.

Mixture Simulator – Results

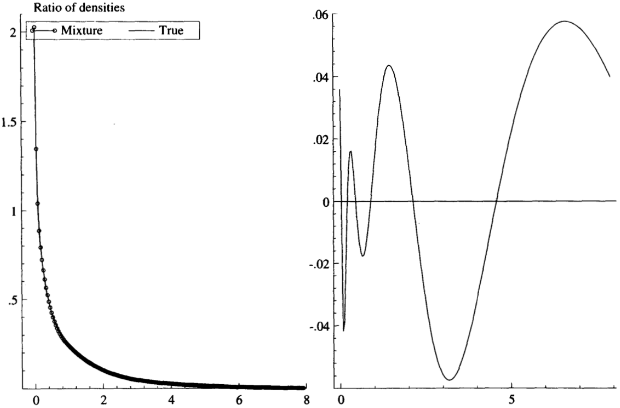

FIGURE 3

Mixture approximation to χ_1^2 density. Left: χ_1^2 density and mixture approximation. Right: the log of the ratio of the χ_1^2 density to the mixture approximation

- The mixture density fits $\log \chi_1^2$ well.

Mixture Simulator – Results

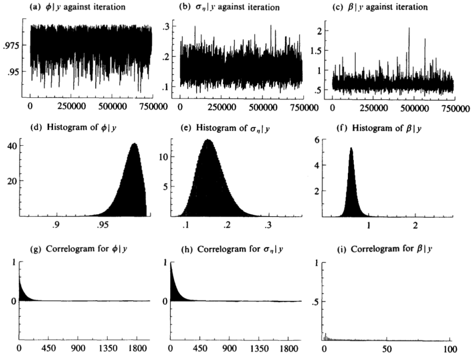

FIGURE 4

Mixture sampler for Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 750,000 iterations were drawn, discarding the first 10,000

- The auto-correlation is improved.
- 20,000 simulations would be sufficient for inferential purposes.

Mixture Simulator – Results

TABLE 5

Daily returns for Sterling against Dollar. Summaries of Figure 2. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 2000, 2000 and 100 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 complete passes of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97779	6·6811e−005	29·776	0·00011093	*−0·690*	*0·203*
$\sigma_\eta \mid y$	0·15850	0·00046128	155·42	−0·00023141	0·0010131	*−0·127*
$\beta \mid y$	0·64733	0·00024217	4·3264	0·00021441	−0·00040659	0·010031
Time	15,374	2·0498				

- Inefficiency is improved.
- Monte Carlo S.E. is reduced.

Integration sampler

Improvement on the mixture sampler

The mixture sampler improves the correlation behaviour of the simulations, the gain is not very big as there is a great deal of correlation between the volatilities and parameters.

- They use the Gaussian structure of $y^*|s, \phi, \sigma_\eta^2$ and sample the joint distribution $\pi(\phi, \sigma_\eta^2, h, \mu)$ by sampling (ϕ, σ_η^2) and (h, μ) independently to overcome this.
- Evaluate the likelihood $f(y^*|s, \phi, \sigma_\eta^2)$ using an augmented version of the Kalman filter to integrate out μ and h, similar to the previous filtered joint normal, the augmented version gives joint normal (See Appendix 2):

$$(\beta, c_1 + Z_1\gamma_1, ..., c_n + Z_n\gamma_n)|y, \theta \tag{17}$$

where $\beta \equiv \mu|y^*, s, \phi, \sigma_\eta^2 \sim \mathbb{N}(\tilde{\mu}, \sigma_{\tilde{\mu}}^2)$

Integration sampler

- We have the posterior $\pi(\phi, \sigma_\eta^2 | y^*, s)$:

$$\pi(\phi, \sigma_\eta^2 | y^*, s) \propto \pi(\phi)\pi(\sigma_\eta^2) f(y^* | s, \phi, \sigma_\eta^2) = \pi(\phi)\pi(\sigma_\eta^2) \frac{f(y^* | s, \phi, \sigma_\eta^2, \mu = 0)\pi(\mu = 0)}{\pi(\mu = 0 | y^*, s, \phi, \sigma_\eta^2)}$$

$$\propto \pi(\phi)\pi(\sigma_\eta^2) \prod_{t=1}^n F_t^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=1}^n v_t^2 / F_t\right) \exp\left(\frac{1}{2\sigma_{\hat{\mu}}^2 \tilde{\mu}^2} \tilde{\mu}^2\right) \sigma_{\hat{\mu}},$$

- Regarding the distribution, we can sample this by drawing from some densities $g(\phi, \sigma_\eta^2)$ and accept using the M-H probability of move:

$$\min\left\{1, \frac{\pi(\phi^{(i)}, \sigma_\eta^{2(i)} | y^*, s)}{\pi(\phi^{(i-1)}, \sigma_\eta^{2(i-1)} | y^*, s)} \frac{g((\phi^{(i-1)}, \sigma_\eta^{2(i-1))}.}{g(\phi^{(i)}, \sigma_\eta^{2(i)})}\right\}.$$

Integration sampler

Integration Sampler

1. Initialize $s, \phi, {\sigma_\eta}^2$, and μ
2. Sample $(\phi, {\sigma_\eta}^2)$ from $\pi(\phi, {\sigma_\eta}^2 | y^*, s)$ using a M-H suggestion based on $g({\sigma_\eta}^2, \phi)$, accepting with the M-H probability
3. Sample $h, \mu | y^*, s, \phi, {\sigma_\eta}^2$ using the augmented simulation smoother given in the Appendix 2
4. Sample $s | y^*, h$ as in the previous algorithm
5. Go to 2

- Can jointly draw (h, μ) and proposal density $g(\cdot)$ is free

Integration sampler – Results

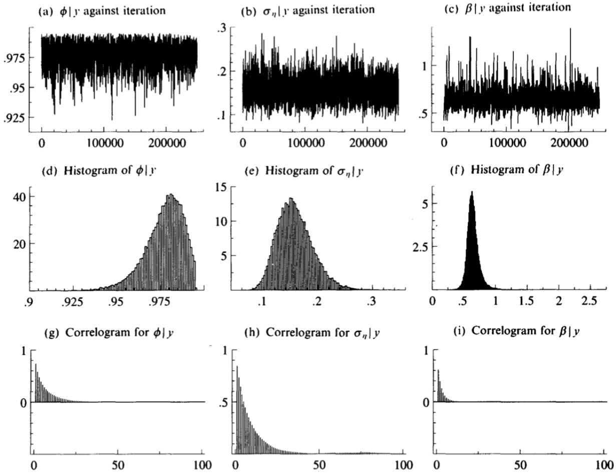

FIGURE 5

The integration sampler for Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 250,000 iterations were drawn, discarding the first 250

Integration sampler – Results

TABLE 6

Daily returns for Sterling against Dollar. Summaries of Figure 5. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 100, 100 and 100 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 complete passes of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97780	6·7031e−005	9·9396	0·00011297	*−0·699*	*0·205*
$\sigma_\eta \mid y$	0·15832	0·00025965	16·160	−0·00023990	0·0010426	*−0·131*
$\beta \mid y$	0·64767	0·00023753	1·4072	0·00021840	−0·00042465	0·010020
Time	8635·2	3·4541				

- Inefficiency is further improved.
- Auto-correlation is further improved.
- These suggest that 2000 samples from this generator would be sufficient.

Re-weighting

Improvements on approximation errors

$$y^* = h_t + z_t, \text{where } z_t \text{ is not } \log \chi_1^2$$

The mixture approximation produces minor approximation error, so they proposed this re-weighting step for correction.

- Write the mixture approximation as making draws from $k(\theta, h|y^*)$, define

$$\begin{aligned} w(\theta, h) &= logf(\theta, h|y) - logk(\theta, h|y^*) \\ &= const + logf(y|h) - logk(y^*|h), \end{aligned} \tag{18}$$

where $f(y|h) = \prod_{t=1}^{n} f_N\{y_t|0, exp(h_t)\}$

and $k(y^*|h) = \prod_{i=1}^{n} \prod_{i=1}^{k} q_i f_N(y_t{}^*|h_t + m_i - 1.2704, v_i{}^2)$

Re-weighting

- Both $f(\cdot)$ and $k(\cdot)$ are Gaussian, then

$$\begin{aligned} Eg(\theta)|y &= \int g(\theta)f(\theta|y)d\theta \\ &= \frac{\int g(\theta)\exp\left[w(\theta,h)\right]k(\theta,h|y^*)d\theta dh}{\int \exp\left[w(\theta,h)\right]k(\theta,h|y^*)d\theta dh} \end{aligned} \tag{19}$$

- So we can re-weighting the MCMC draws according to

$$Eg(\hat{\theta})|y = \sum_j g(\theta^j)c^j \tag{20}$$

where $c^j = \exp\left[w(\theta^j, h^j)\right] / \sum_i \exp\left[w(\theta^i, h^i)\right]$

Re-weighting

- As the mixture approximation is very good, we would expect that the weights c^j would have a small variance.
- The log-weights are close to being normally distributed with a standard deviation of around 1

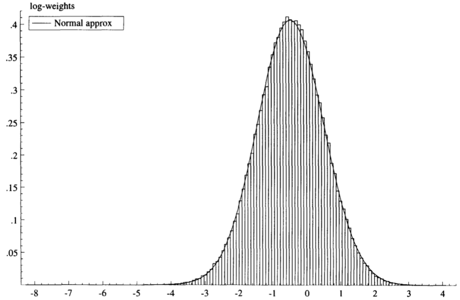

FIGURE 6
Histogram of the log of the $M \times c^j$ for 250,000 sweeps for the integration sampler and a corresponding approximating normal density with fitted mean and standard deviation. All the weights around zero would indicate a perfect sampler

- 250,000 samples from Figure 5 is reweighted.
- The Monte Carlo precision has improved dramatically(?) See the comparison:

TABLE 6

Daily returns for Sterling against Dollar. Summaries of Figure 5. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 100, 100 and 100 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 complete passes of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97780	6·7031e−005	9·9396	0·00011297	*−0·699*	*0·205*
$\sigma_\eta \mid y$	0·15832	0·00025965	16·160	−0·00023990	0·0010426	*−0·131*
$\beta \mid y$	0·64767	0·00023753	1·4072	0·00021840	−0·00042465	0·010020
Time	8635·2	3·4541				

TABLE 7

Daily returns for Sterling against Dollar. Summaries of reweighted sample of 250,000 sweeps of the integration sampler. The Monte Carlo S.E. of simulation is computed using a block one tenth of the size of the simulation. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 complete passes of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97752	7·0324e−005	11·20	0·00010973	*−0·685*	*0·204*
$\sigma_\eta \mid y$	0·15815	0·00024573	14·81	−0·00022232	0·00096037	*−0·129*
$\beta \mid y$	0·64909	0·00025713	1·64	0·00021181	−0·00039768	0·0098312
Time	10,105	4·0423				

Remaining Challenges : Model Diagnosis and Comparison

Remaining Challenges

- How to diagnoses the model?
- How to compare MCMC estimated models?

Solution

Provide a unified set of tools for a complete analysis of SV models, including

- estimation
- likelihood evaluation
- filtering
- diagnostics for model failure
- computation of statistics for comparing non-nested models

Particle Filter under the Bayesian World

Recall that

$$\begin{aligned} y_t &= \beta e^{h_t/2}\varepsilon_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \end{aligned} \tag{21}$$

- The objective is to obtain a sample of draws from $h_t|Y_t, \Theta$ given a sample of draws h^1_{t-1},...,h^M_{t-1} from $h_{t-1}|Y_{t-1}$. Such a algorithm is called a particle filter
- From Bayes theorem,

$$f(h_t|Y_t, \Theta) \propto f(y_t|h_t, \Theta) f(h_t|Y_{t-1}, \theta) \tag{22}$$

Particle Filter

$$f(h_t|Y_t,\Theta) \propto f(y_t|h_t,\Theta)f(h_t|Y_{t-1},\theta) \tag{23}$$

where

$$f(h_t|Y_{t-1},\theta) = \int f(h_t|h_{t-1},\theta)f(h_{t-1}|Y_{t-1},\theta)\,dx \tag{24}$$

- The former $f(h_t|h_{t-1},\theta)$ is the pdf of a normal distrubtion with mean $\mu + \phi(h_{t-1} - \mu)$ and variance σ_η^2
- The latter $f(h_{t-1}|Y_{t-1},\theta)$ estimated from the sample
- $f(h_{t-1}|Y_{t-1},\theta) \simeq \frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j,\theta)$

A-R Sampling for Particle Filter

$$f(h_t|Y_t,\theta) \propto f(y_t|h_t,\theta)\frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j,\theta) \tag{25}$$

How to Sample the Above Distribution?

- importance sampling
- an efficient accept-reject sampling

After some algebra,

$$f^*(y_t,h_t,\theta)\frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j,\theta) \leq g^*(h_t,h_{t|t-1},\theta)\frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j,\theta) \tag{26}$$

These suggest a accept-reject procedure for drawing h_t The probability of acceptance is $f^*(y_t,h_t,\theta)/g^*(h_t,h_{t|t-1},\theta)$

Filtered Results

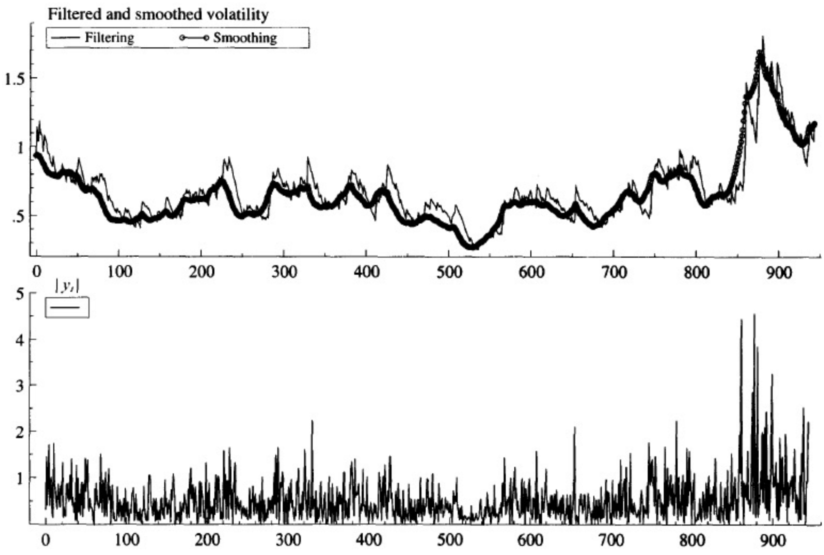

FIGURE 7

Model Diagnostics

Decompose Prediction Density

We decompose the prediction density into

$$f(y_{t+1}|Y_t,\theta) \leq \int f(y_{t+1}|Y_t,h_{t+1},\theta) f(h_{t+1}|Y_t,h_t,\theta) f(h_t|Y_t,\theta)\, dh_{t+1}\, dh_t \tag{27}$$

- We sample the density use the method of composition
- $f(h_t|Y_t,\theta)$ comes from the filtered algorithm. Other two parts are normal density.

Model Diagnostics

From Model Check to Normality Check

$$Pr(y_{t+1}^2 \leq y_{t+1}^{o2} | Y_t, \theta) \simeq u_{t+1}^M \tag{28}$$

- We borrow Rosenblatt (1952)'s idea.
- If the model is correctly specified, u_{t+1}^M converges in distribution to iid uniform random variables as M goes to infinity.

Model Diagnostics

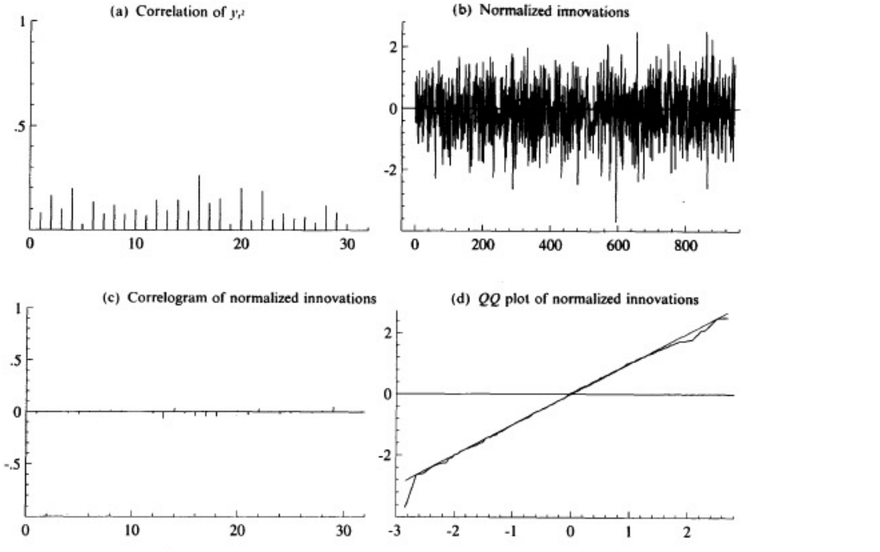

FIGURE 8

Diagnostic checks. Graph (a): correlogram of y_t^2. Graph (b): normalized innovations. Graph (c): the corresponding correlogram. Graph (d): associated QQ-plot.

Model Comparison

GARCH Model

- GARCH model is commonly used in the literature.
- We compare the SV model with a Gaussion GARCH and a Student t GARCH.
- The model setting for the Gaussian GARCH(1,1) model is

$$\begin{aligned} y_t | Y_{t-1} &\sim \mathbb{N}(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \end{aligned} \tag{29}$$

Statistics for Comparison

- Likelihood ratio
- Bayes factors

Model Comparison - Likelihood Ratio

Likelihood Ratios Statistics

$$LR_y = 2logf(y|M_1, theta_1 - logf(y|M_0, theta_0) \quad (30)$$

- The log likelihood is not analytically derived, but instead, sampled using draws from the filtering simulator.

TABLE 12

Non-nested LR tests of the SV model against the ARCH models. In each case the 99 simulations were added to the observed LR_y to form the histograms. The reported r-th rankings are the r-th largest of the observed LR test out of the 100 LR_y tests conducted under SV or GARCH model

	SV versus GARCH			SV against t-GARCH		
Series	Observed	Rank SV	Rank GARCH	Observed	Rank SV	Rank GARCH
Sterling	19·14	81st	100th	−2·68	29th	79th
DM	11·00	61st	100th	−3·84	9th	87th
Yen	19·84	99th	100th	−30·50	1st	1st
SwizF	53·12	100th	100th	−3·62	20th	98th

Model Comparison - Likelihood Ratio

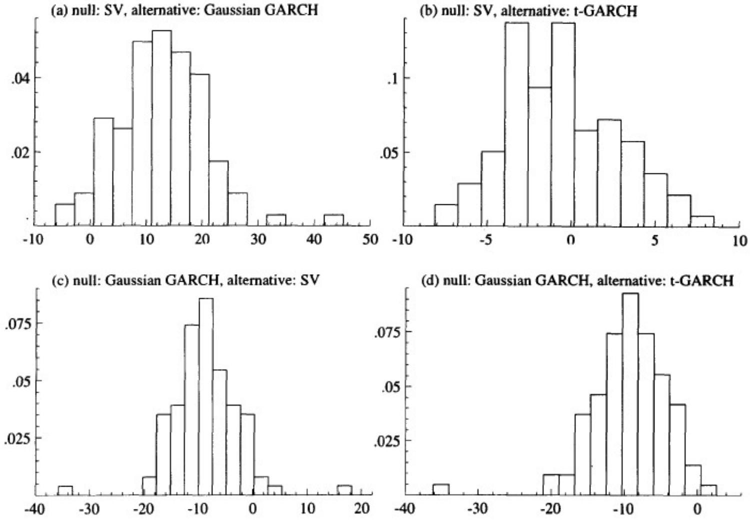

FIGURE 9

Non-nested testing. Graphs (a)–(b) LR_y computed when SV is true. Graph (a): SV against a GARCH model. Graph (b): SV against a t-GARCH. The observed values are 19·14 and −2·68 respectively, which are 80th and 29th out of the 100 samples. Graphs (c)–(d): LR_y computed when GARCH model is true. Graph (c): GARCH

Model Comparison - Bayes Factor

Bayes factors

$$\log f(y|M_1) - \log f(y|M_0) = \\ (\log f(y|M_1,\theta_1) + \log f(\theta_1|M_1) - \log f(\theta_1|M_1,y)) - \\ (\log f(y|M_0,\theta_0) + \log f(\theta_0|M_0) - \log f(\theta_1|M_0,y)) \tag{31}$$

TABLE 12

Non-nested LR tests of the SV model against the ARCH models. In each case the 99 simulations were added to the observed LR_y to form the histograms. The reported r-th rankings are the r-th largest of the observed LR test out of the 100 LR_y tests conducted under SV or GARCH model

	SV versus GARCH			SV against t-GARCH		
Series	Observed	Rank SV	Rank GARCH	Observed	Rank SV	Rank GARCH
Sterling	19·14	81st	100th	−2·68	29th	79th
DM	11·00	61st	100th	−3·84	9th	87th
Yen	19·84	99th	100th	−30·50	1st	1st
SwizF	53·12	100th	100th	−3·62	20th	98th

SV model extensions

1.More complicated dynamics

Three ways of performing Bayesian analysis of the SV model: single move, offset mixture and integration sampling. A useful framework:

$$h_t = c_t + Z_t\gamma_t, where \gamma_{t+1} = d_t + T_t\gamma_t + H_t u_t \tag{32}$$

2.Missing observations

Generalize to any amount of missing data because this argument show that SV model holding at a much finer discretization than the observed data.

3.Heavy-tailed SV models

Help overcome the comparative lack of fit to use an ad hoc scaled Student t distribution. $\varepsilon_t = \sqrt{\frac{\upsilon-2}{\upsilon}}\,\zeta_t / \sqrt{\chi^2_{t,\upsilon}/\upsilon}, \quad \text{where} \quad \zeta_t \overset{iid}{\sim} \mathcal{N}(0,1),\ \chi^2_{t,\upsilon} \overset{iid}{\sim} \chi^2_{\upsilon}$

SV model extensions

4.Semi-parametric SV

Semi-parametric density estimation along with MCMC type algorithms for the updating of the mixture parameters.

5.Prior sensitivity

Easily modified to assess the consequences of changing the prior. One can reweight the simulation to overcome the bias caused by the offset mixture.

6.Multivariate factor SV models

Use MCMC methods on a simplified version by exploiting the conditional independence structure of the multivariate model to allow the repeated use of univariate MCMC methods.

The basis of the N dimensional factor SV model will be

$$y_t = Bf_t + \varepsilon_t,$$

where

$$\begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} \sim \mathcal{N} \langle 0, \operatorname{diag} \{\exp(h_{1t}), \ldots, \exp(h_{Nt}), \exp(h_{N+1t}), \ldots, \exp(h_{N+Kt})\} \rangle,$$

where f_t is K dimensional and

$$(h_{t+1} - \mu) = \begin{pmatrix} \phi_\varepsilon & 0 \\ 0 & \phi_f \end{pmatrix} (h_t - \mu) + \eta_t, \qquad \eta_t \sim \mathcal{N} \left\{ 0, \begin{pmatrix} \Sigma_{\varepsilon\eta} & 0 \\ 0 & \Sigma_{f\eta} \end{pmatrix} \right\}.$$

Conclusion

- A variety of new simulation-based strategies for estimating stochastic volatility(SV) models.
- A comparison of the SV model in raletion to the popular heavy tailed version of GARCH model is provided for the first time.
- Develop an interesting set of methods for filtering the volatilities and obtaining diagnostics for model adequacy.

Thank You!