Inference for Lévy Driven Stochastic Volatility Models via Sequential Monte Carlo

Ajay Jasra

ajay.jasra@imperial.ac.uk

Department of Mathematics

Imperial College London

Talk given as part of the ‘Workshop on Bayesian Inference’, ISM, 21st August

Outline

- Introduction
- Black Scholes Model and the Need for Stochastic Volatility
- A Lévy Driven Model
- Markov chain Monte Carlo (MCMC) Methods
- Sequential Monte Carlo (SMC) Methods via the Expected Auxiliary Variable principle (EAV)
- Analysis of Standard and Poors (S & P) 500 Data
- Summary

Introduction

- In recent years stochastic volatility models have become very important in finance; for example to price options, predict share prices and for hedging. See for example: Heston (1993), Duffie et al. (2000) and Barndorff-Nielsen & Shephard (2001).
- The models incorporate complicated stochastic processes (diffusion and jump-diffusion) to describe the evolution of an asset price and subsequent quantities of interest (e.g. arbitrage price process) may be computed by forwards simulation from the model.
- Instead, we will focus upon the Bayesian model estimation problem; to compute estimates of the model parameters in complex scenarios. This precludes the calculation of financial objects of interest (e.g. call options, greeks etc).
- This problem is not always easy; standard inference tools do not always work well.
- We focus on constructing efficient simulation methods for SV models.

The Black-Scholes Model

- One of the most basic models in option pricing is the famous Black-Scholes model (Black & Scholes, 1973). This models the dynamics of an asset price, S_t, as

$$dS_t = S_t\{\mu dt + \sigma dW_t\}$$

with $t \in [0, T]$ and $\{W_t\}$ a standard Wiener process.

- Standard objects of interest can often be computed in closed form, for example a European call option:

$$\mathbb{E}[(S(T) - K)_+]$$

with $(\cdot)_+ = \max\{\cdot, 0\}$ and K the strike price.

- The model assumes that

$$Y_{t,s} \sim \mathcal{N}\Big((\mu - \sigma^2/2)(t-s), \sigma^2(t-s)\Big)$$

with $Y_{t,s} = \log\{S_t/S_{t-1}\}$, $0 < s < t < T$ and $\mathcal{N}(a, b)$ the normal distribution of mean a and variance b.

The Need for Stochastic Volatility

- The problem of the Black-Scholes model is that the log returns $Y_{t,s}$ exhibit non-normal behaviour, with heavy tails, excess skewness and kurtosis (e.g. Bardorff-Nielsen & Shephard (2001)).
- In addition, there is empirical evidence that there are high and low periods of volatility (σ, called volatility clustering) and a leverage effect (anti-symmetric negative correlation between log price and volatility). Such aspects are not represented in the Black-Scholes model.
- As a result, one way to improve the model (albeit leading to an incomplete market model, in most cases) is to introduce a stochastic volatility. That is $\{\sigma_t\}$ is now a stochastic process.
- In such cases $\{\sigma_t\}$ is often represented by Lévy type processes, which allow flexible modelling of the marginal laws and movements of the volatility (e.g. Carr et al. (2003)).

A Lévy Driven Model

- We will focus upon the following model (the Heston model with a variance Gamma model in the log price (Heston, 1993; Madan et al. 1998))

$$\begin{aligned} dX_t &= \mu dt + \sqrt{\sigma_t}[\rho dW_t^1 + \sqrt{1-\rho^2} dW_t^2] + dZ_t \\ d\sigma_t &= \kappa(\nu - \sigma_t)dt + \sigma_V \sqrt{\sigma}_t dW_t^2 \\ dZ_t &= \gamma dG_t + \sigma_Z \sqrt{dG_t} dB_t \end{aligned}$$

  where $X_t = \log(S_t)$, $\{W_t^1\}$, $\{W_t^2\}$ and $\{B_t\}$ are independent Brownian motions and $\{G_t\}$ is a Gamma process (a Lévy process in cts time with Gamma transition densities). The model is termed the stochastic volatility variance gamma model (SVVG).

- The model incorporates a time changed Brownian motion in the movements of the asset price (that is a Brownian motion at times determined by a Gamma process) and a leverage effect (ρ).

- The model features no jumps in the volatility; as stated by Eraker et al. (2003) this is not appropriate for real financial data. However, Li et al (2007) establish, empirically, that this feature is not required for the above model.

A Lévy Driven Model: Bayesian Formulation

- We are interested in parameter estimation for the SVVG model. To do this we take a Bayesian approach and fit models via MCMC.
- Let $x_{t_0:t_n} = (x_{t_0}, \ldots, x_{t_n})$ be the observed log prices and write the (exact) posterior:

$$\begin{aligned}\pi(\theta, \eta | x_{t_0:t_n}) \quad \propto \quad & \prod_{i=1}^{n} [\phi(y_{t_i, t_{i-1}}; \mu(t_i - t_{i-1}) + (z_{t_i} - z_{t_{i-1}}) + \rho \int \sqrt{\sigma_u} dW_u^1 \\ & , (1-\rho^2) \int_{t_{i-1}}^{t_i} \sigma_u du) p\Big(\int_{t_{i-1}}^{t_i} \sigma_u du | \sigma_{t_i}, \sigma_{t_{i-1}}, \theta \Big) \times \\ & p(\sigma_{t_i} | \sigma_{t_{i-1}}, \theta) p(z_{t_i} - z_{t_i - 1}, \theta)] p(\theta)\end{aligned}$$

  where $\phi(u; a, b)$ is the normal density of mean a and variance b, $\theta = (\rho, \mu, \sigma_V, \sigma_Z, \gamma)$ are the parameters and η are all the other unknowns.
- All densities are known pointwise upto a constant (either analytically or numerically.)

A Lévy Driven Model: Bayesian Formulation

- In most complex SV models the transition densities and likelihood are not known exactly. Therefore, we will design our simulation methods for the Euler approximated model, and then we can investigate the accuracy of the simulation and inference in the context of the exact model (work in progress).

- The approximated model is

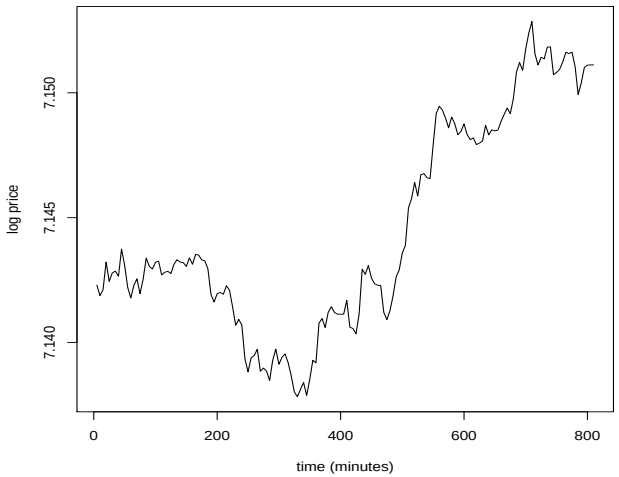
$$\begin{aligned} X_{t+\delta} &= X_t + \mu\delta + \sqrt{|\sigma_t|}\{\rho W^1_{t+\delta,t} + \sqrt{1-\rho^2}W^2_{t+\delta,t}\} + Z_{t+\delta,t} \\ \sigma_{t+\delta} &= \sigma_t + \kappa(\nu - \sigma_t)\delta + \sigma_V\sqrt{|\sigma_t|}W^2_{t+\delta,t} \\ Z_{t+\delta,t} &= \gamma\{G_{t+\delta} - G_t\} + \sigma_J\sqrt{\{G_{t+\delta} - G_t\}}B^\delta_t \end{aligned}$$

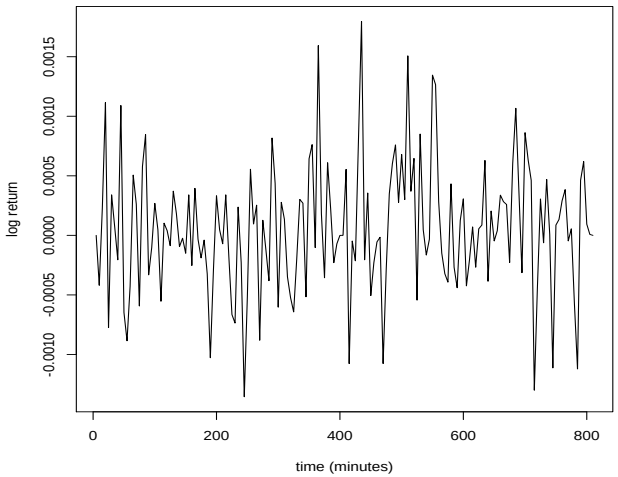
  with $\delta > 0$.

- It is well known (Roberts and Stramer (2001)) that the finer the time-discretization (i.e. as $\delta \downarrow 0$), the better the approximation to the continuous time process, but the worse the performance of MCMC algorithms.

Some MCMC Simulations

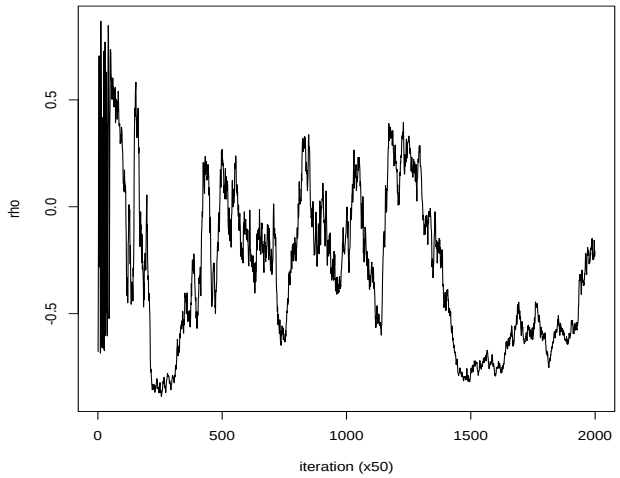
- To motivate the work in the remainder of the talk, I now present some MCMC simulations.
- Recall that MCMC methods simulate from the posterior distribution of interest, by using a Markov transition of stationary distribution that is exactly the posterior.
- We consider 5 minute data from the S & P 500 share index; we look at the data from 13^{th} to 14^{th} February 2006 ($n = 162$).





Some MCMC Simulations

- Below are the sampled ρ from 100000 iterations of an MCMC algorithm thinned to every 50^{th} sample.
- The plot displays extremely high autocorrelations, indicating very poor mixing of the MCMC scheme.
- This problem cannot be solved by increasing the run-time of the MCMC algorithm; reduced serial autocorrelation is not enough to indicate that the algorithm is able to successfully explore the support of the target density.



Some MCMC Simulations

- It may be possible to construct tailored MCMC algorithms that can sample from the target more efficiently, however, there are some drawbacks to this approach:

    1. Such a procedure requires a significant amount of expertise, both in the statistical model, and in simulation techniques.

    2. Such a procedure will require problem specific code, and is may not always work well.

- Our objective is to produce generic methodology, which does not require significant user input and benefits from some of the most advanced simulation techniques.

SMC Methods

- Two simulation methods that we will use are SMC sampler (Del Moral et al. 2006) and EAV (Andrieu et al. 2007) methods.
- SMC samplers is a stochastic simulation technique which simulates from a sequence of densities $\{\widetilde{\pi}_k\}_{1\leq k\leq p}$ defined upon a common space.
- At time k of the algorithm we simulate M samples $\{W_k^{(i)}, X_k^{(i)}\}_{1\leq i\leq M}$ which approximate $\widetilde{\pi}_k$.
- In our case the $X_k^{(i)}$ are M copies of the state simulated in the MCMC scheme (not log prices) and the $W_k^{(i)}$ are importance weights.
- Consider the sequence of densities:

$$\widetilde{\pi}_k(\theta,\eta|x_{t_0:t_n}) \propto L(x_{t_0:t_n};\theta,\eta)^{\zeta_k} p(\eta|\theta)p(\theta)$$

with $0 \leq \zeta_1 < \cdots < \zeta_p = 1$ and $L(\cdot)$ is the likelihood. The idea is to simulate from a simple density $\widetilde{\pi}_1$ and move the samples to sample from π (the posterior).

- To move the samples, we use an MCMC kernel of invariant distribution $\widetilde{\pi}_k$; denote this $P_{k,\tilde{\theta}_k}$, where $\tilde{\Theta} \subseteq \mathbb{R}^q$, is a parameter of the kernel. In this case, the algorithm is set-up so that, for $k \geq 2$;

$$W_k^{(i)} \propto L(x_{t_0:t_n}; \theta_{k-1}, \eta_{k-1})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}.$$

- The (basic) algorithm is then the following:

  (1) At time $k = 1$, for $i = 1, \ldots, M$ sample $X_1^{(i)} \sim \widetilde{\pi}_0$ (an initial distribution) and compute

$$W_1^{(i)} \propto \widetilde{\pi}_1(x_1^{(i)}) / \widetilde{\pi}_0(x_1^{(i)}).$$

  (2) At time $k = 2, \ldots, p$, for $i = 1, \ldots, M$ sample $X_k^{(i)} \sim P_{k,\tilde{\theta_k}}$ and compute

$$W_k^{(i)} \propto L(x_{t_0:t_n}; \theta_{k-1}^{(i)}, \eta_{k-1}^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}.$$

- We will develop methods to set $(\zeta_{1:p}, \tilde{\theta}_{1:p})$ with little user input.

Adaptive SMC Methods

- To set the $\zeta_{1:p}$ we use the following procedure. A criterion which helps to ascertain the performance of SMC methods is the effective sample size (Liu, 2001):

$$ESS_k = \frac{\left(\sum_{j=1}^M W_k^{(i)}\right)^2}{\sum_{j=1}^M \left(W_k^{(j)}\right)^2} = \frac{\left(\sum_{j=1}^M L(x_{t_0:t_n}; \theta_{k-1}^{(i)}, \eta_{k-1}^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}\right)^2}{\sum_{j=1}^M \left(L(x_{t_0:t_n}; \theta_{k-1}^{(j)}, \eta_{k-1}^{(j)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(j)}\right)^2}.$$

  That is, at time k of the algorithm, we may set ESS_k to a prespecified value, $\overline{ESS}_k$, and determine ζ_k, *before* any simulation is performed. $ESS_k - \overline{ESS}_k = 0$ can easily be solved numerically.

- To set the $\tilde{\theta}_k$ we use the following strategy. We approximate at time $k-1$ the mean and then variance associated to $\widetilde{\pi}_{k-1}$, and use this variance in the proposal density; see Andrieu & Moulines (2006) for details.

- The motivation is that if $\widetilde{\pi}_k$ is similar to $\widetilde{\pi}_{k-1}$ (which is required for efficient algorithms), then the variance estimated at time $k-1$ will provide a sensible scaling at time k. Furthermore, we expect to be able to accurately estimate the means and variances at time $k = 1$.

EAV Methods

- Let $\mu : \Theta \to \mathbb{R}^+$ be any unnormalized probability density function of interest, and that it is *not* known pointwise; for example, we might have that:

$$\mu(\theta) = \int_{\Xi} f(\eta, \theta) g(\eta, \theta) d\eta$$

  for some integrable $f : \Theta \times \Xi \to \mathbb{R}^+$, $g : \Xi \times \Theta \to \mathbb{R}^+$, the latter a probability for any $\theta \in \Theta$.

- The idea of the EAV approach is to construct a function $\psi_\theta : \Theta \times \Xi \to \mathbb{R}$, such that we are able to obtain an unbiased estimate of $\mu(\theta)$, (i.e. $\mu(\theta) = \mathbb{E}[\psi_\theta(\eta)]$) for the above example, we could have:

$$\psi_\theta(\eta) = f(x, \eta).$$

- Simulating from the augmented density $f(\eta, \theta) g(x, \eta)$ (up to a normalizing constant) provides samples from μ; this is the standard auxiliary variable approach.

Using the EAV and SMC Methods

- Recall, in the context of SV models there are two problems in the simulation approaches:

    1. When using Euler schemes, the MCMC algorithm can become very poorly mixing as the number augmentation points gets large (i.e. as $\delta \downarrow 0$).

    2. The basic MCMC methods do not always work well.

- We will propose two ideas for dealing with these problems.

- In the context of the first point, we use the EAV idea to integrate out the Euler points; this can lead to better mixing MCMC algorithms.

- In the context of the second point, we use adaptive SMC methods to further improve the simulations.

Using the EAV for SV

- Suppose that θ is only of interest. An efficient simulation scheme, would target the marginal posterior density $\pi(\theta|x_{t_0:t_n})$:

$$\begin{aligned}\pi(\theta|x_{t_0:t_n}) &\propto p(\theta)\int\prod_{i=1}^{n}\left\{\phi(u_{m+1}^{t_i};u_m^{t_i}+b(u_m^{t_i},\theta)\delta,\alpha(u_m^{t_i},\theta)^2\delta)\right\}\times \\ &\quad g(u_{1:m}^{t_1},\ldots,u_{1:m}^{t_n},\theta)d(u_{1:m}^{t_1},\ldots,u_{1:m}^{t_n}) \qquad (1)\\ g(u_{1:m}^{t_1:t_n},\theta) &= \prod_{i=1}^{n}\left\{\prod_{j=1}^{m}\phi(u_j^{t_i};u_{j-1}^{t_i}+b(u_{j-1}^{t_i},\theta)\delta,\alpha(u_{j-1}^{t_i},\theta)^2\delta)\right\}\end{aligned}$$

where $u=(x,\sigma)$ and the drift (resp. diffusion) term has been denoted $b(\cdot)$ (resp. $\alpha(\cdot)$).

- We can apply the EAV framework, and sample from:

$$\pi((u_{1:m}^{t_1:t_m})^{1:N}, \theta | x_{t_0:t_n}) \quad \propto \quad \left\{ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{n} \{\phi(u_{m+1}^{t_i}; (u_m^{t_i})^k + b((u_m^{t_i})^k, \theta)\delta, \alpha((u_m^{t_i})^k, \theta)^2 \delta)\} \right\} \prod_{k=1}^{N} g((u_{1:m}^{t_1:t_m})^k)$$

  which admits the marginal (1) *for any* N. For convergence results on this class of algorithm, see Andrieu & Roberts (2007).

- More transparently if we want to sample

$$\mu(\theta) = \int_{\Xi} f(\eta, \theta) g(\eta, \theta) d\eta$$

  we can sample from

$$\mu(\theta, \eta_{1:N}) \propto \left\{ \frac{1}{N} \sum_{i=1}^{N} f(\eta_i, \theta) \right\} \prod_{i=1}^{N} g(\eta_i, \theta).$$

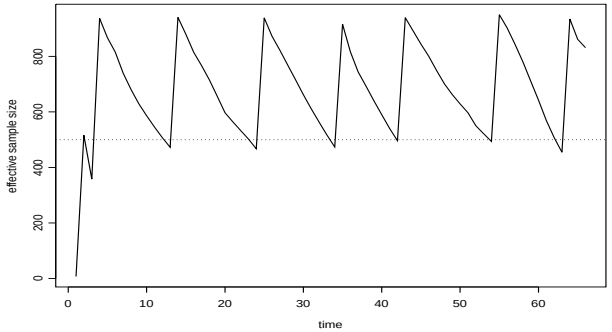
- The potential advantages of sampling (1) over the joint distribution are:
  - In heuristic way, the likelihood is now a mixture over large number of potential latent variables. If sampling from the joint is slowly mixing, this can be compensated by the mixture form of the likelihood. That is, due to the large number of variables in the (latent) state-space, we may expect to represent, simultaneously, many explanations of the data and induce better mixing over θ.
  - In a theoretical way the algorithm can inherit the fast mixing properties of the marginal algorithm; see Andrieu & Roberts (2007).

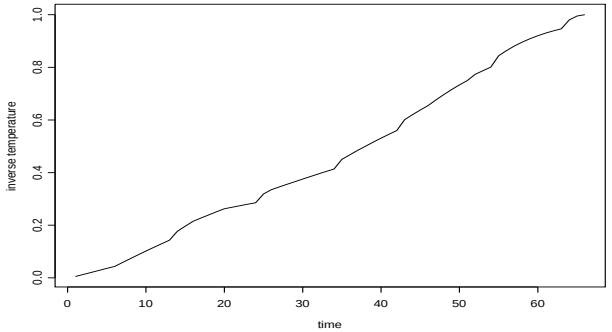
SMC and EAV for SV

- To summarize, we use adaptive SMC methods to sample from the EAV approximated marginal above. This allows us to perform full Bayesian inference for the SV model, in a computationally efficient way.
- The algorithm is then the following:
  - At time $k = 1$, set ζ_1, for $i = 1, \ldots, M$ sample $X_1^{(i)} \sim \widetilde{\pi}_0$ and compute $W_1^{(i)}$. Set $\tilde{\theta}_2$ for the kernel P.
  - At time $k = 2, \ldots, p$, set ζ_k, for $i = 1, \ldots, M$ sample $X_k^{(i)} \sim P_{k,\tilde{\theta}_k}$ and compute $W_k^{(i)}$. If $k < p$, set $\tilde{\theta}_{k+1}$ for each kernel P.
- The convergence properties of this algorithm will be studied in future work.

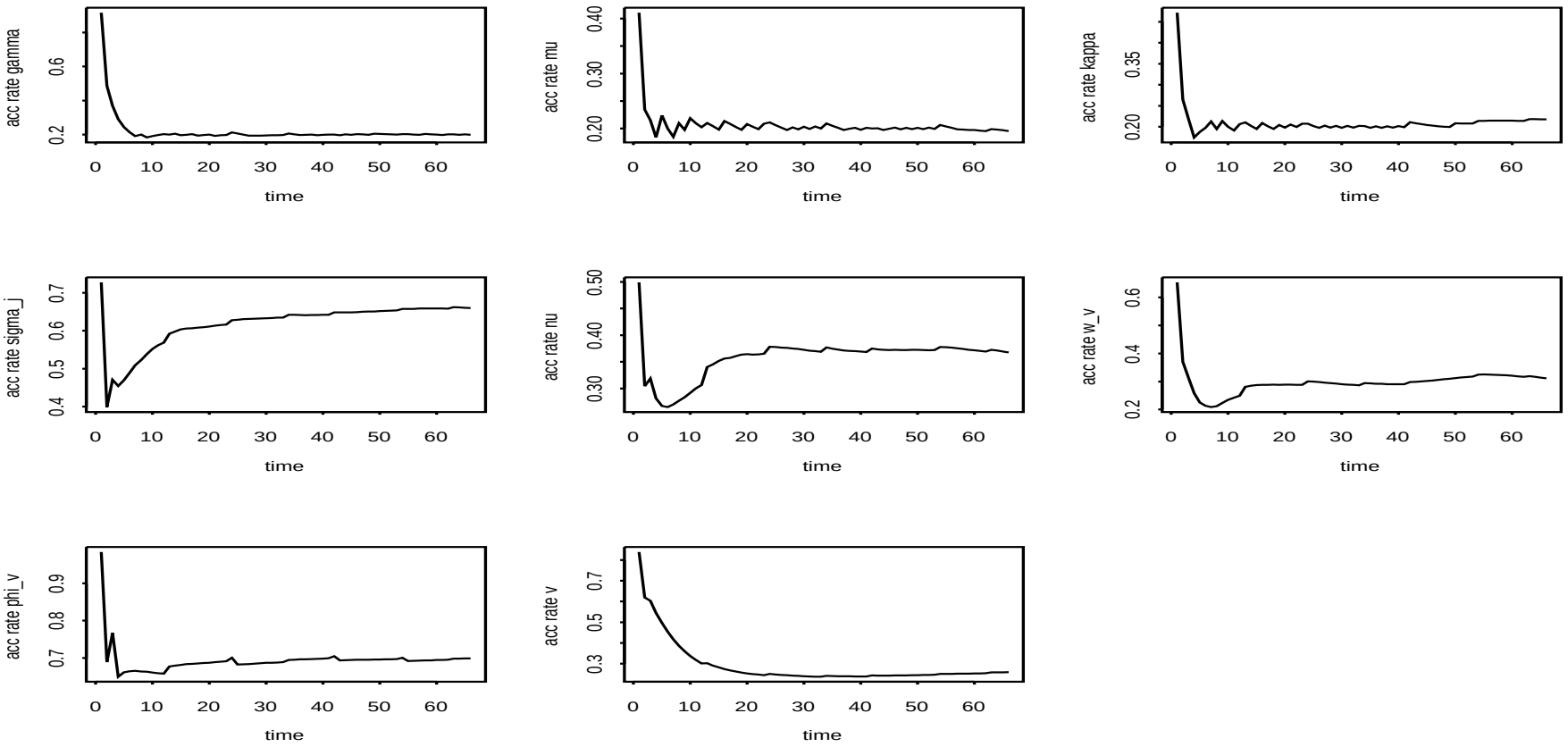
A Demonstration of the Methodology

- We simulate $n = 50$ points and use $N = 5$, $M = 1000$,$\text{E}\bar{\text{S}}\text{S}_k = 0.95\text{ESS}_{k-1}$, $\zeta_1 = 0.005$ and $\widetilde{\pi}_0$ is the prior.
- The algorithm runs so that $p = 65$: this can be controlled via $E\bar{S}S_k$; if it is 'large' then the algorithm takes a long time to run and vice-versa.
- In terms of weight degeneracy, the algorithm performs well - we do not resample the particles irregularly.
- The evolution of $\zeta_{1:p}$ follows the empirical observations of some authors; the temperatures evolve slowly at the start of the algorithm.



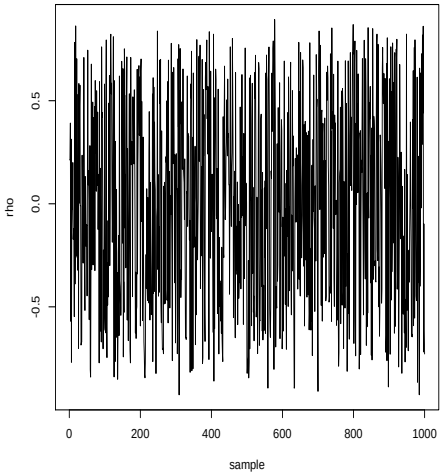


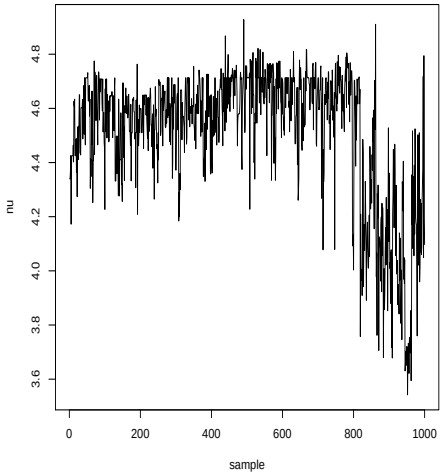
- Below we can see the acceptance rates for the M-H moves; we can see that they are quite reasonable - there are no excessively high or low values.

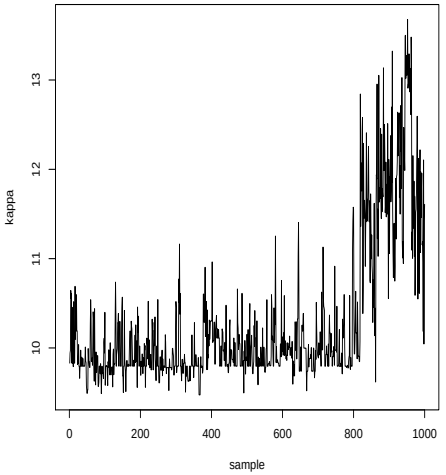


S & P Data Revisited

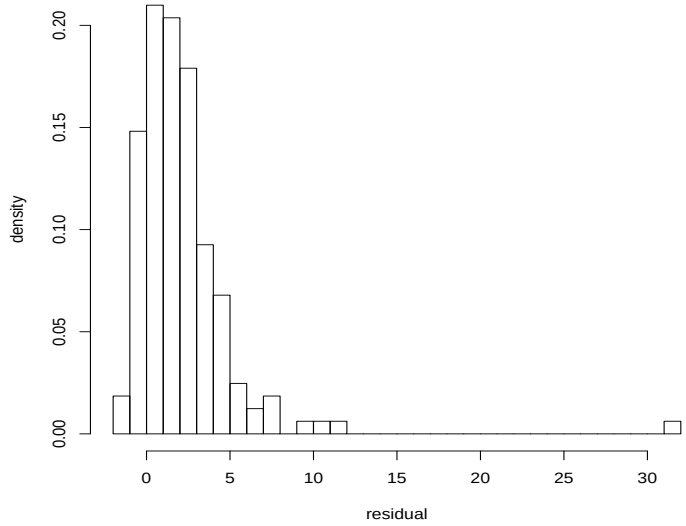
- We return to the S & P 500 data.
- The SMC sampler for $N = 5$ is able to discover multimodal behavior in the marginal posteriors for ρ, ν and κ which did not appear to be present from the output of the MCMC algorithm (for κ and ν the data is not shown).







- Below we can assess the model fit. It can be seen that the model fits the data extremely poorly with the residuals exhibiting an extreme departure from normality; the residuals are positively skewed and indicate, for high frequency data that this model is not always appropriate. We note that we have found for similar subsets of this data that similar poor fit is evident. Whilst we cannot claim that this model is always inappropriate for high frequency data, it does not seem to fit well in these specific cases.



Summary

- In this talk we have discussed simulation methodology for SV models.
- We introduced an adaptive SMC method coupled with the EAV methods to sample from the posterior. It performed better than the MCMC method.
- Our method does not require significant user input, and can work in complex problems, where other methods may not work that well.
- We saw that the SVVG model does not always fit high frequency price data well.

Acknowledgements

- This is a joint work with
- David Stephens (McGill University, CA)
- Arnaud Doucet (UBC, CA).

REFERENCES

ANDRIEU, C. & MOULINES É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Prob.*, **16**, 1462–1505.

ANDRIEU, C., & ROBERTS, G. O. (2007). The pseudo-marginal approach for efficient Monte Carlo computation. Technical Report, University of Bristol.

ANDRIEU, C., DOUCET, A. & ROBERTS, G. O. (2007). The expected auxiliary variable method for Monte Carlo simulation. Technical Report, University of Bristol.

BARNDORFF-NIELSEN, O. E. & SHEPHARD, N. (2001). Non-Gaussian OU-based models and some of their uses in financial economics (with discussion). *J. R. Statist. Soc. B*, **63**, 167–241.

BLACK, F., & SCHOLES, M. (1973). The pricing of corporate liabilites. *J. Political Econm.*, **81**, 637–654.

CARR, P., GEMAN, S., MADAN,D. B¿ & YOR, M. (2003). Stochastic volatility for Lévy processes. *Math. Finance*, **14**, 345–382.

DEL MORAL, P., DOUCET, A. & JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

DUFFIE, D., PAN, J., & SINGLETON, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, **68**, 1343–1376.

ERAKER, B., JOHANNES, M., & POLSON, N. (2003). The impact of jumps in volatilities and returns. *J. Finance*, **58**, 1269–1300.

HESTON, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev Fin* **6**, 327-343.

LI, H., WELLS, M. T. & YU, C. (2007). A Bayesian analysis of return dynamics with Lévy jumps, *Rev. Finan. Studies* (to appear).

MADAN, D. B., CARR, P., & CHANG, E. (1998). The variance gamma process and option pricing. *Eur Finance Rev.*, **2**, 79–105.

ROBERTS, G. O. & STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603-621.

ROBERTS, G. O., PAPASPILIOPOULOS, O. & DELLAPORTAS, P. (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *J. R. Statist. Soc. B*, **66**, 369-393.