# Inference for Lévy Driven Stochastic Volatility Models Via Adaptive Sequential Monte Carlo

Ajay Jasra[*] David A. Stephens[†] Arnaud Doucet[‡]

Theodoros Tsagaris[§]

May 19, 2008

**Abstract**

In the following paper we investigate simulation and inference for a class Lévy driven stochastic volatility (SV) models. The model is comprised of a Heston type model ([22]) with an independent, additive, variance-Gamma process ([28]) in the returns equation. The infinite activity nature of the driving gamma process can capture the observed behaviour of many financial time series, and a discretized version, fit in a Bayesian manner, has been found to be very useful for modelling equity data. In this paper we investigate two aspects associated to this model, and one associated to SV models in general. Firstly, we demonstrate that it is possible to draw exact inference, in the sense of no time-discretization error, from the Bayesian SV model, by the usage of simple results in stochastic calculus and by an auxiliary variable representation of the posterior. Secondly, we investigate the effectiveness of the model in capturing the leverage effect in high-frequency (S & P 500) returns data. Finally, to facilitate the first two points, we introduce a fully automated sequential Monte Carlo (SMC) algorithm, which substantially improves over the standard Markov chain Monte Carlo (MCMC) methods in the literature.

*Key Words*: Stochastic Volatility, Sequential Monte Carlo, Leverage Effect.

## 1 Introduction

Continuous-time stochastic volatility models have become an important way to describe the behaviour of moderate and high frequency financial data. This is due to the inability of the

[*]Institute of Statistical Mathematics{ajay@ism.ac.jp}

[†]McGill University{dstephens@math.mcgill.ca}

[‡]University of British Columbia{arnaud@cs.ubc.ca}

[§]Imperial College London{t.tsagaris@ic.ac.uk}

standard Black-Scholes model [5] to account for returns data that is heavy tailed, with excess kurtosis and skewness and exhibits volatility clustering; all stylized behaviour of such data. In the latter model the movements of the log-price of an asset is modelled by a diffusion process of constant volatility, and it is this aspect of the model which is unsatisfactory. As a result, stochastic volatility (SV) models, with volatility following a stochastic differential equation, have been introduced (e.g. [23]) and used for realistic solutions to problems such as option pricing and hedging. The objective of this article is to investigate the estimation and inference associated to such models, in particular the stochastic volatility variance Gamma (SVVG) model described below.

## 1.1 Continuous Time Modelling of Financial Processes

Following some of initial development of SV models (e.g. [23]) there have been a wide variety of continuous-time models proposed in the literature. We detail some of the approaches here.

In the statistical literature, and particularly from the Bayesian perspective, much of the recent work has stemmed from [4]. This paper introduces a framework for stochastic volatility models, with the volatility following a non-Gaussian Ornstein-Uhlenbeck (OU) process; notable references include, among others: [33] who fit SV models with the changes in the log price modelled via a driftless diffusion and volatilities through a super-position of (OU) processes with Gamma marginals (see also [12] for sequential inference for this model); [21] who, in addition, investigate models with jumps in the returns, an incorporation of a leverage effect and different risk premium coefficients for different component volatilities. In more recent work, [19] look at such OU models when the marginal of the OU process is generalized inverse Gaussian and inverse gamma and introduce an approximation to fractional Brownian motion in the returns equation; in addition [20] generalize the autocorrelation structure of the volatilities to model long-memory properties. The finance literature is slightly more varied, with an investigation of jumps and diffusion in both the volatility and returns; [15] consider exactly this scenario, with both the returns and volatilities following affine jump diffusion (AJD) equations; the volatilities and returns follow a correlated Brownian model with additive jumps; [16] state that the presence of jumps in the volatilities as well as returns is important for modelling real financial data. Interestingly, when both stock and option data are analyzed jointly, [17] shows that 'economically significant performance enhancements' can be yielded. In addition, an investigation of the leverage effect, via high frequency data, and its implication on the volatility structure in SV models has been considered in [6]. In a more recent paper [26] consider the impact and comparison (with affine

jump diffusion models) of models which include, in the jumps on the returns, infinite activity Lévy processes (e.g. [3]).

## 1.2 Observation and Volatility Models Driven by Lévy Processes

In this paper, the class of models that are considered are of the following form:

$$\begin{aligned} dY_t &= \mu dt + \sqrt{V_t}[\rho dW_{1t} + \sqrt{(1-\rho^2)} dW_{2t}] + dZ_t \\ dV_t &= \kappa(\nu - V_t)dt + \sigma_V \sqrt{V_t} dW_{1t} \\ dZ_t &= \gamma dG_t + \sigma_J \sqrt{dG_t} dB_t \end{aligned} \tag{1.1}$$

where $t \in [0,T]$, $\{Y_t\}_{t\in[0,T]}$ are log price data, $\{V_t\}_{t\in[0,T]}$ is the volatility, $\{W_{1t}\}_{t\in[0,T]}$, $\{W_{2t}\}_{t\in[0,T]}$ and $\{B_t\}_{t\in[0,T]}$ are three independent, one dimensional, Brownian motions, $\{G_t\}_{t\in[0,T]}$ is a Gamma process with parameter $\alpha$, so that, for all $t_0, t$,

$$G_{t+t_0} - G_{t_0} \sim \mathcal{G}a(t/\alpha, 1/\alpha)$$

with $\mathcal{G}a$ denoting the Gamma distribution. Also $\rho \in [-1,1]$, the correlation between the instantaneous returns and volatility, $\gamma, \sigma_V, \sigma_J \in \mathbb{R}^+$, $\nu, \mu, \kappa \in \mathbb{R}$ are possibly unknown parameters; we denote the unknown parameters as $\theta$, with corresponding state-space $\Theta$. This model incorporates, via the additive Lévy process in the returns equation, a variance-gamma model ([28]; see also [9]). This is a time-changed or subordinated infinite activity Lévy process, specifically a Brownian motion realized at times governed by a Gamma process (see, for example [3] for more details); such a process allows a model to capture micro structure in trading activity.

The reasons for focussing upon this model are related to two aspects. Firstly, in the paper of [26], they demonstrate that (from a Bayesian perspective) the model out-performs, in terms of fit, some of the most sophisticated AJD models (when considering 20 years of daily returns data from the S& P 500 share index). However, there are two points of interest here; it is possible to fit this model exactly, that is, with no time discretization error, and (as noted above) that [16] state that there is strong evidence to include jump terms in the volatility and returns (also that e.g. [6] show that it is sensible have multiple factors in the volatility). The question is then, is the infinite activity nature of the additive Lévy process able to compensate for the lack of jumps (and one factor) in the volatility? This is an important question, as if such a model can reasonably capture the dynamics of the data, then, in realistic scenarios, it will be easier to (say) approximate and/price options than two factor stochastic volatility models. Here the factors reflect persistent and fast moving volatility structures. Secondly, is an investigation of

the ability of the model to capture the leverage effect. It is known that one of the best ways to study the leverage effect is, unsurprisingly, via looking at high frequency returns data (e.g. [6]). Thus, in this context we will investigate this financially important aspect for SVVG model, given an efficient estimation procedure.

## 1.3 Bayesian Computational Inference

Bayesian SV models are conceptually attractive, however, it is often a complex task to fit them. In the case where the model cannot be fitted exactly in continuous time (that is, Euler discretizations are adopted) there is a vast literature concerning the difficulties of MCMC inference; see [33] and more recently [25]. When it is possible to perform exact inference for SV models, the problems are still very challenging. For example, the work of [32], [21] and [19] all require advanced MCMC techniques and reparameterizations. Indeed, in the case of the first publication a relatively generalized reparameterization scheme, *non-centered* parameterizations ([30]) are adopted. However, despite these improvements, it is well-known that MCMC can often fail to work in high dimensional problems (e.g. [24]).

In this paper we develop a fully automated and adaptive simulation methods based upon the SMC sampler ([11]) approach. Briefly, it requires only the specification of a relatively simple MCMC algorithm and can proceed to produce a highly effective simulation scheme that will often work better than MCMC approaches. Whilst such methods are perhaps too complex to be used in real-life problems (e.g. for automated trading - methods such as in [29] might be used), it is important and enlightening to ascertain the ability of the model to replicate the market when fitted exactly.

## 1.4 Objectives and Structure

To summarize, the objectives of this paper are

- To introduce efficient simulation techniques for SV models in general.
- Given the reliability of these methods, to investigate some inferential aspects of the SVVG model.

This paper is structured as follows; in Section 2 we discuss the SVVG model in more depth with the an auxiliary variable approach outlined; in Section 3 we introduce our data and highlight the simulation difficulties associated to MCMC; in Section 4 our new simulation methods are introduced and investigated, in Section 5 we return to our data and investigate the leverage

effect in more detail and in Section 6 we conclude the paper. Our basic MCMC algorithm is given in the appendix.

# 2 The Variance Gamma Model

The SVVG model detailed in Section 1.2 has been developed as a possible competitor to the widely used AJD models. The model describes the log price through both jump and diffusion components and, in addition, the volatility through a diffusion. As stated in the introduction and investigated by [16], the absence of a jump component in the volatility may not be adequate to model real financial data. That is, jumps in the volatility provide a longer term effect on the returns, than simply a jump component in the returns themselves. [26], however, show that the infinite activity Lévy process model is in fact superior to AJD models (for the examples considered there).

Another key component of the SVVG model is the modelling of leverage (see [21] and [34] for recent discussions in a Bayesian modelling context). Leverage encapsulates the concept that the sign of changes in the price of the underlying $\exp\{Y_t\}$ has an asymmetric affect in contemporaneous or subsequent changes in volatility (or, the causality is *reversed*, that is, there is a volatility feedback effect; see [6] and the references therein). The SVVG model captures leverage through the parameter $\rho$, and received wisdom suggests that $\rho$ is negative in sign, so that instantaneous increases in $Y_t$ are associated with instantaneous decreases in $V_t$, and vice versa.

We now introduce a Bayesian model, and an associated probability density (to the posterior) which allows us to perform inference from the model. Before we continue, we give some notation that is used in the paper.

## 2.1 Notation

For convenience, it is assumed throughout that data are observed at regular time intervals on $[0, T]$; that is $0 < t_1 < \cdots < t_n = T$ (note that our approach can be applied in the irregular case). The following notation is adopted, for any $n \in \mathbb{Z}^+$: for a vector $x_{1:n} := (x_1, \ldots, x_n)$. A measurable space is written $(\mathsf{F}, \mathcal{F})$. If $\Pi$ is a probability measure the expectations of $\Pi-$integrable functions $H$ are denoted $\mathbb{E}_\Pi[H]$. In addition, $\mathbb{T}_n := \{1, \ldots, n\}$ and a generic notation for a probability density is $p(\cdot)$. For a complex number $z$, $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ denotes the real and imaginary parts of $z$ respectively. Given a real-valued function $g$, the sign of $g$ is written $\mathrm{Sgn}(g)$.

## 2.2 Likelihood Construction

In a paper of [8], it is demonstrated that it is possible to simulate exactly from a class of AJD processes. We utilize these ideas, to show that the likelihood for the SVVG model is known point-wise, for discretely observed data. We begin by looking at the volatility process and then the jumps component in the returns. Then given the various random variables associated to these processes, we construct the log price likelihood.

### 2.2.1 The Volatility Process

Inspecting the SDE for the volatility process in equation (1.1), we have for $0 < u$, $t \in (u, T)$,

$$V_t = V_u + \kappa\nu(t-u) - \kappa\int_u^t V_s\, ds + \sigma_V \int_u^t \sqrt{V_s}\, dW_{1s}. \tag{2.2}$$

For each disjoint observation period, we decompose the process into two components corresponding to the two integrals in equation (2.2). For $t > u$, define the random variables

$$V^*_{u,t} = \int_u^t V_s\, ds \qquad\qquad \overline{V}_{u,t} = \int_u^t \sqrt{V_s}\, dW_{1s}$$

with corresponding realized values $v^*_{u,t}$ and $\overline{v}_{u,t}$. We use the decomposition

$$p(v_t, v^*_{u,t}, \overline{v}_{u,t}|\theta, v_u) = p(\overline{v}_{u,t}|v_t, v_u, v^*_{u,t}, \theta)p(v^*_{u,t}|v_t, v_u, \theta)p(v_t|v_u, \theta).$$

The process $\{V_t\}$ follows the Cox-Ingersoll-Ross model, and it is a standard result that the transition density is

$$V_t|v_u, \theta \sim \frac{\sigma_V^2(1-e^{-\kappa(t-u)})}{4\kappa}\chi_d^{'2}\left(\frac{4\kappa e^{-\kappa(t-u)}v_u}{\sigma_V^2(1-e^{-\kappa(t-u)})}\right)$$

where $\chi_d^{'2}(\tau)$ is the non-central chi-squared distribution on $d = 4\kappa\nu/\sigma_V^2$ degrees of freedom and non-centrality parameter $\tau$. Following a result in [31], it is also possible to derive the characteristic function of $V^*_{u,t}$ conditional on $V_t, V_u$ as

$$\begin{aligned}\Phi_{\theta,t,u}(y) \quad = \quad & \frac{\eta(y)e^{-0.5(\eta(y)-\kappa)(t-u)}(1-e^{-\kappa(t-u)})}{\kappa(1-e^{-\eta(y)(t-u)})} \times \\ & \exp\left\{\frac{v_u+v_t}{\sigma_V^2}\left[\frac{\kappa(1+e^{-\kappa(t-u)})}{1-e^{-\kappa(t-u)}} - \frac{\eta(y)(1+e^{-\eta(y)(t-u)})}{1-e^{-\eta(y)(t-u)}}\right]\right\} \times \\ & \frac{I_{0.5d-1}\left[\sqrt{\{}v_tv_u\}\frac{4\eta(y)e^{-0.5\eta(y)(t-u)}}{\sigma_V^2(1-e^{-\eta(y)(t-u)})}\right]}{I_{0.5d-1}\left[\sqrt{\{}v_tv_u\}\frac{4\kappa e^{-0.5\kappa(t-u)}}{\sigma_V^2(1-e^{-\kappa(t-u)})}\right]}\end{aligned}$$

where $\eta(y) = \sqrt{\{}\kappa^2 - 2\sigma_V^2 iy\}$, and $I_d(\cdot)$ is the modified Bessel function of the first kind. Hence the density function is, up to a known constant,

$$p(v^*_{u,t}|v_t, v_u, \theta) \propto \int_{\mathbb{R}^+} \cos(yv^*_{u,t})\mathrm{Re}[\Phi_{\theta,t,u}(y)]dy + \int_{\mathbb{R}^+} \sin(yv^*_{u,t})\mathrm{Im}[\Phi_{\theta,t,u}(y)]dy \tag{2.3}$$

For now, we will ignore the fact that the density is written in terms of an integral - we will demonstrate in Section 2.4.3 that it is possible to draw inference from the correct posterior. Finally, from (2.2), the conditional distribution of $\overline{v}_{u,t}$ is degenerate

$$p(\overline{v}_{u,t}|v_t, v_u, v^*_{u,t}, \theta) = \mathbb{I}_{\{f(v_t, v_u, v^*_{u,t}, \theta)\}}(\overline{v}_{u,t})$$

with

$$f(v_t, v_u, v^*_{u,t}, \theta) = \frac{v_t - v_u - \kappa\nu(t-u) + \kappa\widetilde{v}_{u,t}}{\sigma_V}.$$

From herein the densities $p(\overline{v}_{u,t}|v_t, v_u, \widetilde{v}_{u,t}, \theta)$ are removed from the notation; it is assumed that $\overline{v}_{u,t}$ takes the value $f$ throughout.

### 2.2.2 Variance Gamma Process

In the case of the VG process, we note that the latent Gamma process can be integrated out of the model; due to the independent increments property of Lévy processes, if $Z^*_{u,t} = Z_t - Z_u$ and $G^*_{u,t} = G_t - G_u$ are the increments of $Z_t$ and $G_t$,

$$p(z^*_{u,t}|\theta) = \int p(z^*_{u,t}|g^*_{u,t}, \gamma, \sigma_J)p(g^*_{u,t})\, d(g^*_{u,t})$$

this marginal is known; by Theorem 1 of [28]:

$$p(z^*_{u,t}|\theta) = \frac{2\exp\{\gamma(z^*_{u,t})^2/\sigma\}}{\alpha^{(t-u)/\alpha}\sqrt{2\pi}\Gamma((t-u)/\alpha)}\left(\frac{(z^*_{u,t})^2}{\gamma^2 + 2\sigma^2/\alpha}\right)^{(t-u)/2\alpha - \frac{1}{2}} K_{(t-u)/\alpha - \frac{1}{2}}\left(\frac{\sqrt{(z^*_{u,t})^2(\gamma^2 + 2\sigma^2/\alpha)}}{\sigma^2}\right) \tag{2.4}$$

where $K_c(\cdot)$ is the modified Bessel function of the second kind, $\sigma = \sigma_J\sqrt{t-u}$ and the Gamma process has increment distribution $\mathcal{G}a((t-u)/\alpha, 1/\alpha)$. In all of our simulations $\alpha = (t-u)$ (recall regularly spaced data are observed).

### 2.2.3 The Likelihood

Now, conditionally upon our latent processes, it is straightforward to deduce that the joint density of our observed data is

$$\prod_{i=1}^{n}\left\{\phi(y_{t_i}; y_{t_{i-1}} + \mu(t_i - t_{i-1}) + z^*_i + \rho\overline{v}_i, (1-\rho^2)v^*_i)\right\}$$

where we recall that

$$z^*_i = z^*_{t_{i-1},t_i} \qquad \overline{v}_i = \overline{v}_{t_{i-1},t_i} \qquad v^*_i = v^*_{t_{i-1},t_i}$$

and $\phi(u; a, b)$ is the normal density of mean $a$ and variance $b$ and $t_0 = 0$.

## 2.3 Some Comments on the Model

We remark that if $\rho = 0$ it is straightforward to show that for the discrete-time sampling model, and any $k > 0$

$$\text{Corr}[(Y_{t_i} - Y_{t_{i-1}})^2, (Y_{2t_i - t_{i-1}+k} - Y_{t_i+k})^2] = \frac{\mathbb{C}\text{ov}\big[V_i^*, \int_{t_i+k}^{2t_i-t_{i-1}+k} V_s ds\big]}{\mathbb{V}\text{ar}[V_i^*] + \xi(\alpha, \gamma, \sigma_J, t_i, t_{i-1})}$$

where

$$\xi(\alpha, \gamma, \sigma_J, t_i, t_{i-1}) = \left[\frac{\{\sigma_J \gamma (t_i - t_{i-1})\}^2}{\alpha} + \gamma(t_i - t_{i-1})\right]\left(6\mathbb{E}[V_i^*] - 1\right) + \mathbb{E}[(Z_i^*)^4]$$

and the expectations of the various quantities can be found in [7] and [18]. Indeed, if $\mu = 0$ then

$$\mathbb{C}\text{ov}\left[V_i^*, \int_{t_i+k}^{2t_i-t_{i-1}+k} V_s ds\right] = \exp\{-\kappa k(t_i - t_{i-1})\}(1 - \exp\{-\kappa\delta\})^2 \frac{\nu\sigma_V^2}{2\kappa^3}$$

and the correlation decays exponentially fast. This shows, without the presence of leverage, that the parameters of the VG process contribute to the auto-correlation structure of the squared returns, and will only decrease the correlation from the integrated volatility process; i.e. faster than the original exponential rate. That is to say, that we may not be confident that the addition of the additive VG process will be sufficient to substitute for an additional factor in the volatility.

## 2.4 The Posterior

We now construct the posterior density and in particular, show how to perform inference with respect the appropriate posterior density. This is not simple, as $p(\widetilde{v}_{u,t}|v_t, v_u, \theta)$ in (2.3) is written in terms of an integral.

### 2.4.1 Priors for the Parameters

Now that we have noted that it is possible to write the likelihood of the data and the process, exactly, given the observed data, we now discuss the priors which are used. The prior dependence structure is

$$p(\theta) = p(\mu)p(\gamma)p(\sigma_J)p(\rho)p(\kappa, \sigma_V, \nu).$$

It is assumed that

$$\mu \sim \mathcal{N}(\xi_\mu, \varrho_\mu) \qquad \gamma \sim \mathcal{N}(\xi_\gamma, \varrho_\gamma) \qquad \sigma_J \sim \mathcal{G}a(\tau_{\sigma_J}, \upsilon_{\sigma_J})$$

where $\mathcal{N}(\xi, \kappa)$ is the normal distribution of mean $\xi$ and variance $\varrho$. Also $\rho = 2\overline{\rho} - 1$ with

$$\overline{\rho} \sim \mathcal{B}e(\tau_\rho, \upsilon_\rho)$$

and $\mathcal{B}e(\tau, \upsilon)$ is the Beta distribution. The joint structure of $\kappa, \sigma_V, \nu$ is slightly more intricate; an improper prior is used, for $d_{\max} > 1$

$$p(\kappa, \sigma_V, \nu) = \kappa^{\tau_\kappa - 1} e^{-\upsilon_\kappa \kappa} \mathbb{I}_{\{\kappa : d_{\max} > 0.5d > 1\}}(\kappa) \Gamma_p(\sigma_V; \tau_{\sigma_V}, \upsilon_{\sigma_V}) \Gamma_p(\nu; \tau_\nu, \upsilon_\nu)$$

where $\Gamma_p(\cdot)$ is the Gamma probability density. The specification is made, to avoid any difficulties in the evaluation of the modified Bessel function in the density (2.3). In addition, it also ensures that the Feller condition is satisfied and a stationary distribution of $\{v_t\}$ exists; in the Heston model with no leverage this places restrictions on the moments of the returns (e.g. [18]).

### 2.4.2 Posterior Density of Interest

Putting together the elements above the posterior density function is

$$\pi(\theta, v_{1:n}, v^*_{1:n}, z^*_{1:n} | y_{0:n}) \propto \prod_{i=1}^{n} \left[ p(y_i | y_{i-1}, v^*_i, z^*_i, \theta) p(v^*_i | v_i, v_{i-1}, \theta) p(v_i | v_{i-1}, \theta) p(z^*_i | \theta) \right] p(\theta). \quad (2.5)$$

We assume that $v_0 = 1$, although it is possible to include $V_0$ as an unknown parameter in the Bayesian analysis. We also assume that $y_{t_0}$ is the observed log price at the start of the trading period. In addition, for any collection of data points the normalizing constant is indeed finite; the normal densities are upper bounded (with function dependent upon the $y_{1:n}$) and the resulting factors integrate to a finite value.

If it were possible, we would seek to simulate from the posterior density (2.5). However, due to the integrals in (2.3), it is not possible to evaluate the posterior density pointwise, up-to a normalizing constant; this is necessary in order to perform MCMC simulations. Whilst Fourier inversion methods could be employed, this is computationally expensive (note that [18] use a bivariate normal approximation, in a similar context, but we would like to assess the model when fitted exactly). We now give an auxiliary variable framework, which defines a density on an extended space, which allows us to compute expectations wrt the posterior and simultaneously reduces the computational expense compared to a Fourier inversion.

### 2.4.3 An Auxiliary Variable Approach

The idea of extending the state-space of the target density, such that a marginal is the original target is often used in the Monte Carlo literature; e.g. [11]. In our approach we use a similar idea, except that we rely on the fact that transformed versions of expectations with respect to our new target are equivalent to expectations wrt our posterior of interest. This calculation is essentially based on the construction of an efficient importance sampling density, and is similar to that in [27].

The idea is simply to define the auxiliary densities. Let $G : \Theta \times (\mathbb{R}^+)^2 \times [0,T]^2 \to \mathbb{R}$ be such that

$$G_{\theta,t,u}(v^*, w) = \cos(v_i^* w_i)\mathrm{Re}(\Phi_{\theta,t_i,t_{i-1}}(w_i)) + \sin(v_i^* w_i)\mathrm{Im}(\Phi_{\theta,t_i,t_{i-1}}(w_i)).$$

Then, for $i \in \mathbb{T}_n$, introduce a bounded set $A \subset \mathbb{R}^+ \times \mathbb{R}^+$, and the density

$$\widetilde{p}(v_i^*, w_i | v_i, v_{i-1}, \theta) \propto |G_{\theta,t,u}(v^*, w)|\mathbb{I}_A(v_i^*, w_i) + q_\theta(v_i^*, w_i)\mathbb{I}_{A^c}(v_i^*, w_i) \tag{2.6}$$

where $q_\theta$ is essentially an arbitrary function, henceforth taken as $e^{-v^*-w}|G_{\theta,t,u}(v^*, w)|$. The new target density from which we seek to simulate is then

$$\widetilde{\pi}(\theta, v_{1:n}, v_{1:n}^*, z_{1:n}^*, w_{1:n} | y_{0:n}) \propto \prod_{i=1}^{n} \Big[ p(y_i | y_{i-1}, v_i^*, z_i^*, \theta)\widetilde{p}(v_i^*, w_i | v_i, v_{i-1}, \theta)p(v_i | v_{i-1}, \theta)p(z_i^* | \theta) \Big] p(\theta). \tag{2.7}$$

Define the function $H : \Theta \times (\mathbb{R}^+)^2 \times [0,T]^2 \to \mathbb{R}$

$$H_{\theta,t,u}(v_i^*, w_i) = \frac{G_{\theta,t,u}(v_i^*, w_i)}{\widetilde{p}(v_i^*, w_i | v_i, v_{i-1}, \theta)}.$$

In order to compute expectations of $\pi-$integrable functions $h$, we use the following identity (related to the work of [27])

$$\mathbb{E}_\pi[h(X_n^*)] = \frac{\mathbb{E}_{\widetilde{\pi}}\big[h(X_n^*)\prod_{i=1}^n \big\{H_{\theta,t_i,t_{i-1}}(V_i^*, W_i)\big\}\big]}{\mathbb{E}_{\widetilde{\pi}}\big[\prod_{i=1}^n \big\{H_{\theta,t_i,t_{i-1}}(V_i^*, W_i)\big\}\big]} \tag{2.8}$$

where we have used $x_n^* = (\theta, v_{1:n}, v_{1:n}^*, z_{1:n}^*)$.

The identity (2.8) is a standard importance sampling identity. The motivation for the set $A$ in (2.6) is the fact that the function $H_{\theta,t,u}(v^*, w)$, for fixed $(\theta, t, u)$, is only conditionally integrable on $\mathbb{R}^+ \times \mathbb{R}^+$. Thus we cannot use

$$\widetilde{p}(v_i^*, w_i | v_i, v_{i-1}, \theta) \propto |G_{\theta,t,u}(v_i^*, w_i)|$$

with

$$H_{\theta,t,u}(v^*, w) = \mathrm{Sgn}(G_{\theta,t,u}(v^*, w))$$

(which is the approach in [27]) as the densities $\widetilde{p}$ and $\widetilde{\pi}$ are improper. The simple idea is that if the set $A$ is 'large' then the weights $H_{\theta,t,u}$ will be in the set $\{-1, 1\}$ and the empirical variance of our estimates will not be too large.

It should be noted that the computational cost of using the auxiliary target (2.7) against the original one (2.5), is the fact that for each data, the modified Bessel function (computed using the approach in [1]) need only be computed *once* for each data point; this is quite an expensive calculation, and our problem is too complicated to be able to resort to look-up tables. If a Fourier inversion approach is used, then it is necessary to use the discrete Fourier transform; accepting

that this is reasonably accurate, it still requires more than one evaluation of the modified Bessel function. Clearly, our approach is at the expense of increasing the size of the state-space, however we have found the approach to be computationally much faster. The principal advantage of our proposed method is that it is indeed exact, as the integral in (2.3) is not approximated. The only error in our inference is related to Monte Carlo and numerical round-off error of the computer.

In summary our inference procedure is to simulate from the target distribution (2.7) and use the identity (2.8) to compute approximations of $\mathbb{E}_\pi[h]$.

# 3 An Initial Analysis of High Frequency Data

In the following Section we discuss the initial analysis of some high-frequency S&P 500 5 minute return data. We then illustrate that standard MCMC methods do not work well for the SVVG posterior. This motivates the work in the next Section for automatic SMC methods.

## 3.1 Initial Data Analysis

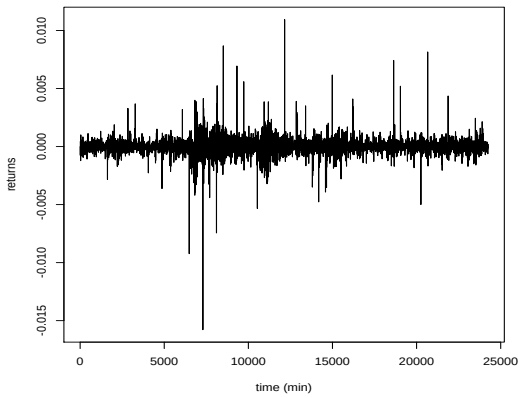
The data we investigate are 5 minute returns data from the S&P 500 share index on the NYSE, recorded between 2nd February 2007 (14:30 BST) and 3rd May 2007 (20:05 BST) to constitute $n = 4829$ data points. The data are presented in Figure 1 and table of summary statistics are given in Table 1.

Figure 1 (a) depicts the (log) returns which exhibit periods of (relatively) high volatility around 7500 minutes and 12500 minutes. In Figure 1 (b) we can observe the the correlation between the absolute returns $|R_{t_i}| = |\log(S_{t_i}) - \log(S_{t_{i-1}})|$ and the lagged returns $R_{t_i - j}$ (i.e. the positive $x$ axis is the correlation between current absolute returns and returns in the past). The plot exhibits the well-known leverage effect. In other words, future volatility (with absolute returns used as a proxy for volatility) are correlated with the returns, but that the reverse correlation is around 0 (i.e. future returns are uncorrelated with the volatility) and that the sign of the correlation is negative. As noted in [6], the spike at zero can be associated to volatility feedback effect, which implies that the causality of the leverage effect is reversed (i.e. the effect of pricing volatility affects the returns as opposed to the returns (and hence financial leverage) affecting volatility).

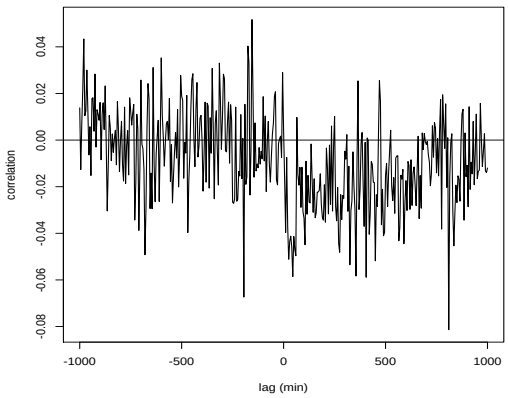
To estimate our model, we will thin the data to hourly returns, and this data is seen in Figure 1 (c) and (d). The returns data is much the same, but the correlation structure is slightly modified, with the volatility feedback effect emphasized. The summary statistics (of the hourly returns)

in Table 1 are reasonably consistent with share data that are observed over longer periods and lower frequency; see e.g. [21].

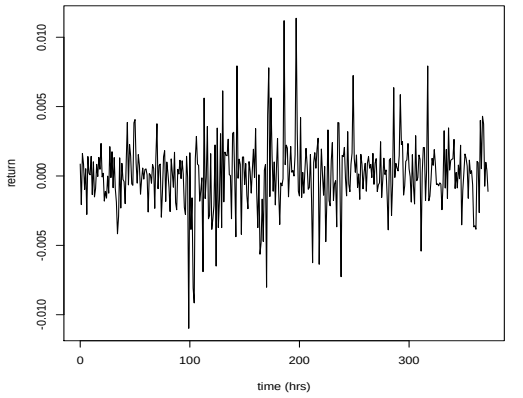
We are now interested to see whether the results of [26] can be confirmed in the context of fitting the model exactly and whether the leverage effects seen here, can be replicated by the model (given well-estimated parameters).



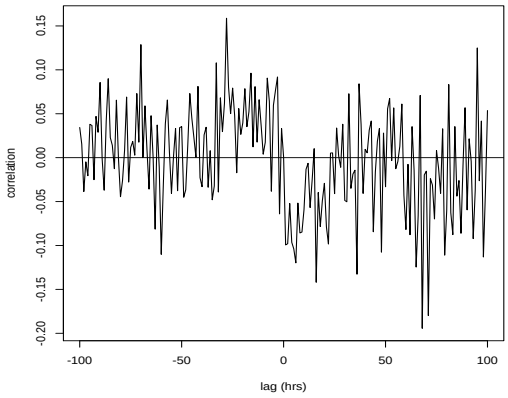

(a) 5 Minute Returns Data.




(b) Correlations; absolute and lagged returns.




(c) Hourly Returns Data.




(d) Correlations; absolute and lagged returns.

Figure 1: The Data and Correlation Structure. In (a) we have 5 minute return data from the S&P 500 share index and in (b) the correlation between the absolute returns $|R_{t_i}| = |\log(S_{t_i}) - \log(S_{t_{i-1}})|$ and the lagged returns $R_{t_i-j}$.

## 3.2 MCMC Simulation

In order to investigate this data, using the SV model in Section 2, we apply the MCMC algorithm in the appendix (and prior settings as in Section 5.1), to simulate from the density (2.7). The

| Data | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|
| $Y_t$ | $8.90\times10^{-5}$ | $6.59\times10^{-6}$ | 0.01 | 6.57 |
| Estimated | -0.22 | 1.38 | -0.98 | 6.52 |

Table 1: Summary statistics from the S&P 500 share index data (373, hourly returns). The skewness and kurtosis are calculated as $\tau_3/\tau_2^{3/2}$ and $\tau_4/\tau_2^2$, with $\tau_i$ the $i^{\text{th}}$ central moment. The estimated row is on the basis of the SMC results in Section 5.

algorithm is run for 100000 iterations with simulation parameters set by prior simulations. The C++ code took approximately 50 minutes on a Pentium 3GHZ machine. In Figure 2 we can observe some results.

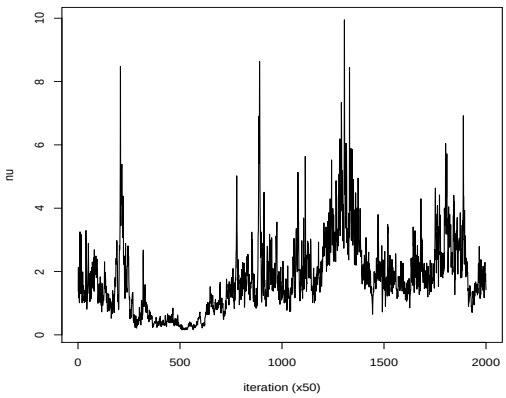
In Figure 2 we can see the sampled value of $\nu$ and the autocorrelations. The plot indicates extremely poor mixing of the MCMC sampler and suggests that the resulting inferences will be very unreliable. As noted in Section 1.3, it possible to construct advanced MCMC methods to deal with the problems seen here. There are some drawbacks with these ideas:

- It will require a significant adaptation of the computer code;
- The methods may require more user input;
- Such methods, due to the nature of MCMC, will not always work well.

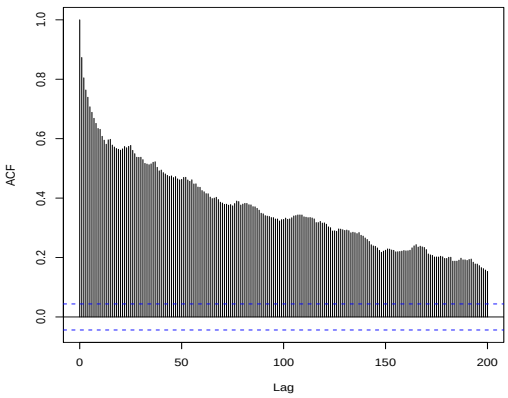
In reference to the first point, reparameterizations will often require a lot of the code to be adapted. This can take a long time and even more so introduce the possibility of bugs in the code (note that C++ programming of the MCMC algorithm already took us a reasonable time to debug). In terms of the second, we may need to (as for this example) put in a large amount of information in terms of simulation parameters - this may be difficult. Lastly, it has been observed in the literature (e.g. [24]) that standard MCMC can be substantially improved upon, using SMC methods. Such approaches will now be introduced and improved so as to deal with the first two aspects above.

# 4 Automatic and Adaptive SMC Methods

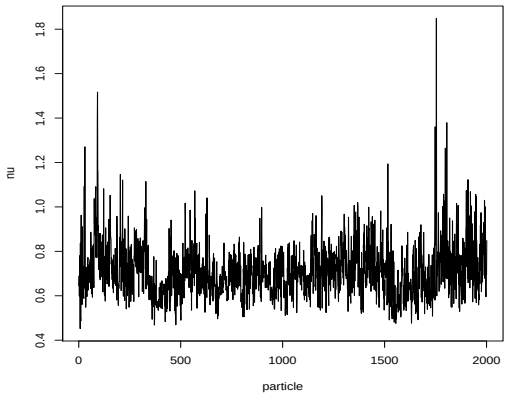
In complex problems, it is well-known that MCMC methods do not always work well; this was seen in in Section 3. Therefore, it is important to develop more advanced simulation techniques; one such method is termed SMC samplers. We introduce a fully automatic SMC sampler and illustrate that it substantially improves over the MCMC simulations in the previous Section.




(a) Sampled $\nu$.




(b) Autocorrelations of $\nu$.




(c) Unweighted Particles, $\nu$.

Figure 2: The sampled $\nu$ and their autocorrelations (a-b) and from the SMC Sampler (c). We ran our MCMC algorithm for 100000 iterations and thinned to every $50^{\text{th}}$ sample. The SMC sampler was run with 2000 particles.

## 4.1 SMC Samplers

An SMC sampler is a stochastic simulation technique that is designed to simulate from a sequence of densities $\{\widetilde{\pi}_k\}_{1\leq k\leq p}$ defined upon a common space $(\mathsf{E}, \mathcal{E})$. The method uses a combination of sequential importance sampling/resampling (e.g. [14]) and MCMC methods to provide an efficient simulation procedure.

At step $k$ of the algorithm, we simulate a cloud of particles

$$\{W_k^{(i)}, X_k^{*(i)}\}_{1\leq i\leq M}$$

which are designed to approximate $\widetilde{\pi}_k$. In our case the $X_k^{*(i)}$ will represent $M$ copies of the state

simulated in the MCMC scheme and the $W_k^{(i)}$ are importance weights which are used to correct for the fact that the sampling densities are not the sequence of target distributions. We remark that whilst this procedure may seem very computationally expensive, it can often be the case that the computational cost is comparable to MCMC; see for example [10].

## 4.2 SMC Sampler for SV

Consider the sequence of densities:

$$\widetilde{\pi}_k(\theta, v_{1:n}, v^*_{1:n}, z^*_{1:n}, w_{1:n}|y_{0:n}) \propto \prod_{i=1}^{n} \left[ p(y_i|y_{i-1}, v_i^*, z_i^*, \theta)^{\zeta/n} \widetilde{p}(v_i^*, w_i|v_i, v_{i-1}, \theta) p(v_i|v_{i-1}, \theta) p(z_i^*|\theta) \right] p(\theta)$$

with $0 \leq \zeta_1 < \cdots < \zeta_p = 1$. The idea is to simulate from an easy to sample density $\widetilde{\pi}_1$ and move the particles to enable us to simulate from the target of interest $\widetilde{\pi}_p = \pi$. To move the particles, in an efficient way, it is often sensible to adopt an MCMC kernel of invariant distribution $\widetilde{\pi}_k$; denote this $P_{k,\tilde{\theta}_k} : (\mathsf{E} \times \tilde{\Theta}) \times \mathcal{E} \to [0,1]$, where $\tilde{\Theta} \subseteq \mathbb{R}^q$, $q \in \mathbb{Z}^+$ is a parameter of the kernel (i.e. differs from the $\Theta$ already introduced). In this case, the algorithm is set-up so that the importance weight, at time $k \geq 2$, is written

$$W_k^{(i)} = L(y_{0:n}; (y^*_{k-1})^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}$$

with

$$L(y_{0:n}; y^*) = \prod_{i=1}^{n} [p(y_i|y_{i-1}, v_i^*, z_i^*, \theta)]$$

and $y^* = (\theta, v^*_{1:n}, z_{1:n})$. The 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

$$W_k^{(i)} \propto L(y_{0:n}; y^*_{k-1})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}.$$

To deal with the problem of weight degeneracy (e.g. [14]), a resampling step may be added.

### 4.2.1 An Adaptive SMC Sampler

One of the drawbacks of the specified model is the fact that the parameters $\{\zeta_k\}$ and $\{\tilde{\theta}_k\}$ have to be set; this is not always straightforward for non-experts.

To set the $\{\zeta_k\}$ we use the following procedure. A criterion which helps to ascertain the performance of SMC methods (and in general any dependent sampling method) is the effective sample size (see e.g. [14], it measures the number of samples relative to an independent simulation approach), defined here as

$$ESS_k = \frac{\left(\sum_{j=1}^M W_k^{(i)}\right)^2}{\sum_{j=1}^M \left(W_k^{(i)}\right)^2} = \frac{\left(\sum_{j=1}^M L(y_{0:n}; (y_{k-1}^*)^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}\right)^2}{\sum_{j=1}^M \left(L(y_{0:n}; (y_{k-1}^*)^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}\right)^2}.$$

That is, at time $k$ of the algorithm, we may set $ESS_k$ to a prespecified value, $\overline{ESS}_k$, and determine $\zeta_k$, *before* any simulation is performed; $ESS_k - \overline{ESS}_k = 0$ will need to be solved numerically; this is a relatively simple task.

To set the $\tilde{\theta}_k$ we use the following strategy. An important area of MCMC methods are adaptive MCMC techniques [2], which seek to update the parameters of an MCMC kernel in the light of the simulated values. In our context, all of the kernels which are parameter dependent, are random-walk Metropolis moves, so for a parameter in $x_k^*$ (the entire state at time $k-1$) we approximate at time $k-1$ the mean and then variance of the marginal, and use this variance in the proposal density; see [2] and references therein for further 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 $k = 1$.

The algorithm is then the following:

- At time $k = 1$, set $\zeta_1$, for $i = 1, \ldots, M$ sample $X_1^{(i)} \sim \widetilde{\pi}_0$ (an initial distribution) and compute
$$W_1^{(i)} \propto \frac{\widetilde{\pi}_1(x_1^{(i)})}{\widetilde{\pi}_0(x_1^{(i)})}.$$
Set $\tilde{\theta}_2$ for each kernel.

- At time $k = 2, \ldots, p$, set $\zeta_k$, for $i = 1, \ldots, M$ sample $X_k^{(i)} \sim P_{k,\tilde{\theta}_k}$ and compute
$$W_k^{(i)} \propto L(y_{0:n}; (y_{k-1}^*)^{(i)})^{\zeta_k - \zeta_{k-1}} W_{k-1}^{(i)}.$$
If $k < p$, set $\tilde{\theta}_{k+1}$ for each kernel.

For the convergence analysis see [13].

# 5 Results for the S&P 500 data

In this Section we return to the S & P 500 data from Section 3. We use the SMC method developed above to sample from the target density (2.7). Note that removing the start and end of the days trading, does not significantly change our conclusions.

## 5.1 Priors

The prior parameters were taken as: $\xi_\mu = \xi_\gamma = 0$, $\varrho_\mu = \varrho_\gamma = 100$, $\tau_\rho = 1/7$, $\upsilon_\rho = 1$, $d_{\max} = 2$, $A = (0, 10000) \times (0, 10000)$ (as in Section 2.4.3), $t_i - t_{i-1} = 1/7$ for each $i \in \mathbb{T}_n$, and all other prior parameters are 1. In terms of sensitivity, we found for short time series (50 data points) that the inferences were quite sensitive to the prior settings for the parameters; this effect erodes for larger data sets.

On the choice of $d_{\max}$, we set it to be 2, as it appeared that the posterior density values were larger for this specification, MCMC algorithm mixed slightly better and the computation times were faster; unfortunately we have traded off inferential aspects for the computational ones here. Note that whilst the inferences changed for larger values of $d_{\max}$, it did not appear that they 'improve' as a result.

For the set $A$, if the set is smaller, then the functions $H_{\theta,t,u}(v^*, w)$ are more variable, and inferences more unreliable. However, once the set is larger (than chosen) that inferences were quite robust to this choice.

## 5.2 SMC Sampler

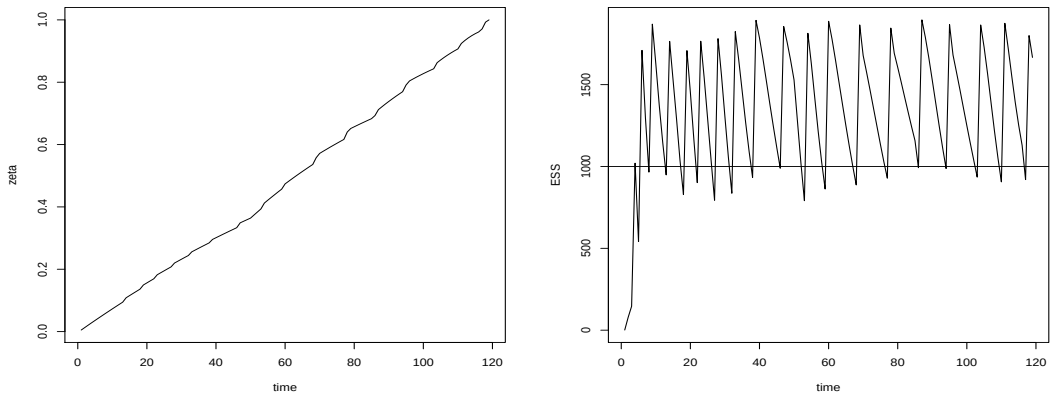
The algorithm is run with $M = 2000$ particles with a resampling threshold of 1000 particles (with systematic resampling; see [14] and the references therein) and $\zeta_1 = 0.005$, $\overline{ESS}_k = 0.95 ESS_{k-1}$. The initial proposal variances were the same as used by the MCMC sampler and the initial distribution of the particles, for the parameters $\theta$ to be the prior densities (that is the same distributional family) at the estimates of the posterior $\tilde{\pi}_1$ and the other values simulated from the priors. To set the inverse temperature we let $\overline{ESS}_k = 0.95 ESS_{k-1}$. In addition, to ensure that the acceptance rates in the Metropolis-Hastings steps did not get too large or small we used the following procedure; if the rate exceeded 0.7 (resp. fell below 0.2) we multiplied the current standard deviation by 5 (resp. 1/5). The CPU time was approximately 110 minutes. The performance of the algorithm can be seen in Figures 3, 4 and 2 (c).

In Figure 3 (a) the evolution of the $\{\zeta_k\}$ can be seen. The algorithm is run for 120 steps; this is controlled by the parameter $\overline{ESS}_k$. If it is made larger, the algorithm will take longer to

run, but is likely to suffer from less weight degeneracy. The Figure shows that the parameter is piece-wise linear function of time, with a slight increase in gradient at $k = 50$. This corresponds to the fact that the particle system has to adapt to the introduction of the likelihood; this is reflected in Figure 3 (b). Here the system of particles finds it difficult to adapt to the target, initially, then as time progresses, they begin to stabilize. The performance, in this respect, is quite good as it does not indicate any serious problems in terms of weight degeneracy.

Figure 4 displays the Metropolis-Hastings acceptance rates. Here we can see that the rates are never too low, although at times some of the rates are a little high. We note that we observed some sensitivity of the rates to initial parameters; if the proposal variances are very 'high' then the algorithm finds it difficult to recover (i.e. has very low acceptance rates). However, reasonable initial values can be determined on the basis of the (poorly mixing) MCMC algorithm.

In Figure 2 (c) the (unweighted) $\nu$ particles can be seen. Here it is apparent that the dependence of the parameters is substantially lower. Despite the fact that the CPU time is higher, we expect the samples to be more representative of the target density (due to the population-based nature of the algorithm). Note that since the particles are not weighted, it is not necessarily the case that the parameters support should overlap (with the MCMC).




(a) $\zeta_k$.

(b) Effective Sample Size.

Figure 3: Performance of the adaptive SMC Sampler. In (a) the evolution of the parameter $\zeta$ is displayed and (b) the effective sample size; the horizontal line is the resampling threshold. We ran our SMC sampler with 2000 particles.

## 5.3 Inference

We now consider the inferences from our model, given the success of the simulation method.

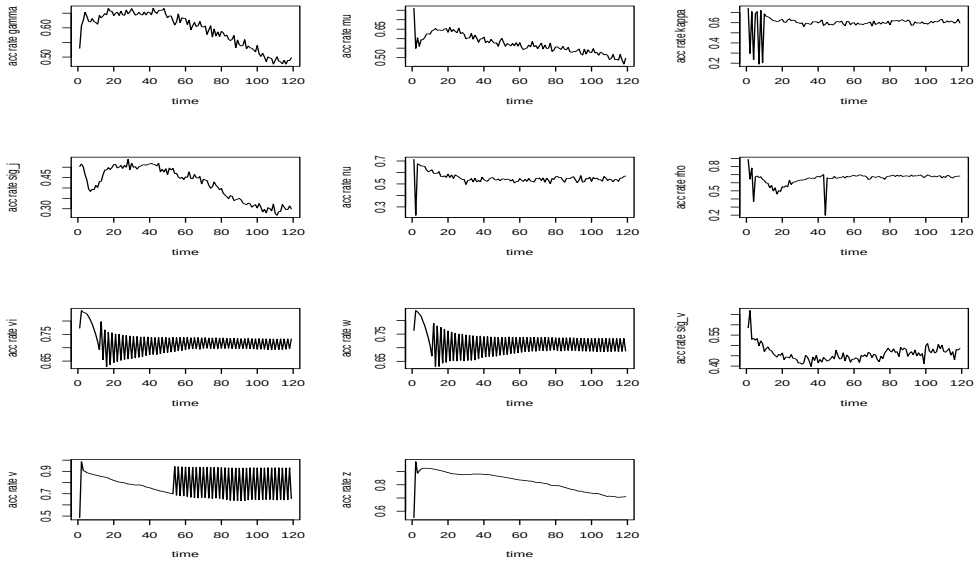



Figure 4: Acceptance Rates of MCMC moves in our SMC Sampler. The acceptance rates for each of the parameters is given. The initial variance is as the MCMC algorithm in Section 3.2.

In Table 1 the (one step ahead) posterior predictive summaries can be seen. This is calculated through an additional (for each particle) 10000 Monte Carlo samples (and Rao Blackwellization). The Table reveals that the model can capture the kurtosis of the data (and to an extent the mean), but that both the variance and skewness are over and under estimated respectively.

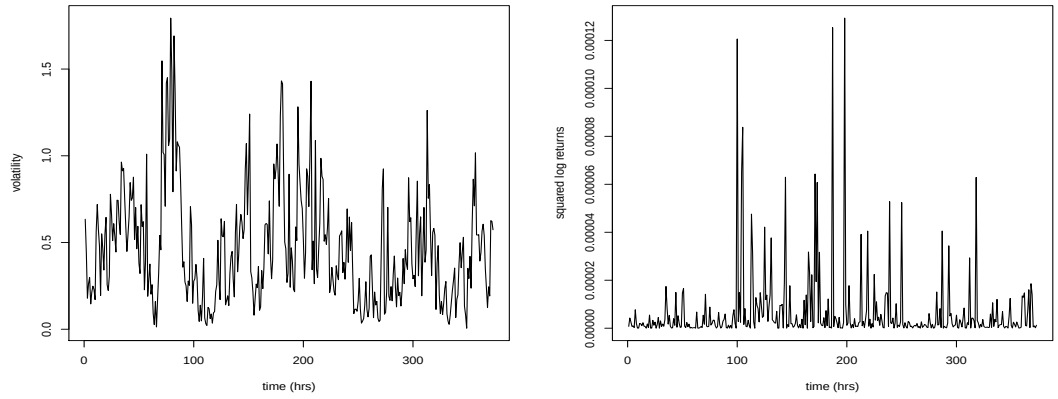
In terms of the parameter estimates (in Table 2) the most noticeable feature, is the relatively small leverage parameter (in comparison to a large number of previous studies). We found that, for this data, this was the case for a wide variety of settings of the priors and other model parameters. The volatility parameters indicate a high level of persistence (the median of $\kappa = 0.86$), and it is likely that, as indicated by the results discussed in the previous paragraph, that a faster moving volatility component is needed fit the data. The point is somewhat reinforced by Figure 5. In the Figure, we can observe that the volatility (a) generally follows the trend of the squared returns (b), but that it is slow to revert to its mean level.

Finally, looking at Figure 6 and comparing to that of Figure 1 (computed as for the predictive moments), we can observe that the general structure is maintained (a drop below 0), but that the return to a zero level appears to be too fast. This coincides with the results of [6], in the context of the Heston model. That is, it does not appear that, via adding a variance Gamma process, this model can always capture the leverage structure present in the data. We added a risk premium to the model, but it did not seem to improve the models ability to capture the

| Parameter | 0.025 | Median | 0.975 |
|---|---|---|---|
| $\kappa$ | 0.42 | 0.86 | 1.08 |
| $\gamma$ | -4.07 | -3.12 | -3.06 |
| $\sigma_J$ | 0.74 | 2.80 | 2.89 |
| $\mu$ | 0.46 | 0.98 | 1.36 |
| $\nu$ | 0.45 | 0.63 | 1.03 |
| $\sigma_V$ | 0.69 | 0.89 | 1.08 |
| $\rho$ | -0.007 | -0.002 | -0.001 |

Table 2: Posterior Medians and 95% Credible Intervals from the S&P 500 share index data (373, hourly returns).

leverage effect.




(a) Estimated Volatility.

(b) Squared Log Returns.

Figure 5: Estimated Volatility and Squared Log Returns from the S&P 500 share index data (373, hourly returns).

# 6 Summary

In this paper we have introduced a general methodology to sample from complex stochastic volatility models as well as investigated inference for the SVVG model when applied to high frequency returns data. We established that our adaptive SMC approach can provide reliable simulation results from models that can be difficult to fit using standard MCMC approaches. We also showed that the SVVG model cannot always replicate the structure that is present in the

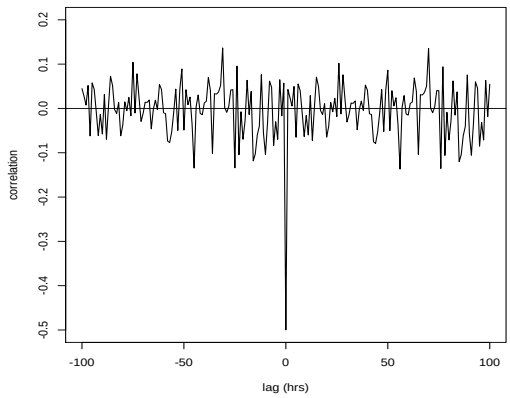



Figure 6: Predictive Correlation between Absolute Returns and Lagged Returns from the S&P 500 share index data (373, hourly returns).

data. That is, the addition of the variance gamma process in the returns is not always sufficient, in the presence of a single volatility factor. We expect that the latter would be required (or the addition of jumps in the volatility), at least, to enable the model to fit well, consistently, although we do not doubt that the model may be useful (due to the results of [26]). It may also suffice to introduce richer leverage structures as in [21]. It is straightforward to adapt our simulation technique to this context.

We remark that whilst the data-set used is not too large, i.e. that more data is perhaps needed to estimate the parameters of the model. However, as noted in the introduction, such data are found in real trading situations and hence the performance of the model is of interest in this context.

The approach detailed in this paper can be applied for any SV model, and combined with any sophisticated MCMC algorithm. The algorithm is adaptive, requiring little user input and raises the possibility of generic computer packages for SV models. In addition, we have shown how to draw Bayesian inference when transition densities are only known through their characteristic function. This pushes the boundaries of the number of SV models which may be fitted (either exactly, or through discretization) and is a useful tool for investigating model inferences.

## Acknowldgement

The main part of this work was done whilst the first author was at Imperial College London. We also thank Jim Griffin for some conversations related to this work.

# Appendix

## MCMC Algorithm

We now detail the MCMC steps that are used in this paper. Recall that we seek to sample from the density (2.7), which is associated to the variables $(\theta, v_{1:n}, v^*_{1:n}, v^*_{1:n}, w_{1:n})$ and $\theta = (\mu, \gamma, \sigma_J, \rho, \kappa, \sigma_V, \nu)$. We will simply state the move types; all acceptance probabilities are suppressed, but easy to calculate.

Our algorithm is a sweep over the variables $(\theta, v_{1:n}, v^*_{1:n}, v^*_{1:n}, w_{1:n})$, and all moves are random-walk Metropolis-Hastings move. The term 'normal random walk Metropolis kernel' is defined to be a Metropolis-Hastings kernel, with a proposal density that is normal, centred at the current state. The moves are

- $\mu$ is updated by a normal random walk Metropolis kernel.
- $\gamma$ is updated by a normal random walk Metropolis kernel.
- $\sigma_J$ is updated by a normal random walk Metropolis kernel on the log scale.
- $\rho$ is updated by a normal random walk Metropolis kernel on the $\log([1+\rho]/[1-\rho])$ scale.
- $\kappa$ is updated by a normal random walk Metropolis kernel on the $\log(\kappa - u^*)$ scale, $u^* = \sigma_V^2/2\nu$.
- $\nu$ is updated by a normal random walk Metropolis kernel on the $\log(\nu - u^*)$ scale, $u^* = \sigma_V^2/2\kappa$.
- $\sigma_V$ is updated by a normal random walk Metropolis kernel on the $\log(\sigma_V/[\sigma_V - u^*])$ scale, $u^* = \sqrt{\{2\kappa\nu\}}$.
- $v_{1:n}$ are updated one-in-turn by a normal random walk Metropolis kernel on the log scale.
- $v^*_{1:n}$ are updated one-in-turn by a normal random walk Metropolis kernel on the log scale.
- $z_{1:n}$ are updated one-in-turn by a normal random walk Metropolis kernel.
- $w_{1:n}$ are updated one-in-turn by a normal random walk Metropolis kernel on the log scale.

The proposal variances were chosen by prior tuning, although as we saw in Section 4 they are set adaptively in the SMC case.

# References

[1] Amos, D. E. (1986). Algorithm 644, a portable package for Bessel functions of a complex argument and non-negative order. *ACM Trans. Math. Software*, **12**, 265–273.

[2] Andrieu, C. & Moulines É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Prob.*, **16**, 1462–1505.

[3] Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge: CUP.

[4] 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.

[5] Black, F., & Scholes, M. (1973). The pricing of corporate liabilites. *J. Political Econm.*, **81**, 637–654.

[6] Bollerslev, T., Litvinova, J. & Tauchen, G. (2006). Leverage and volatility feedback effects in high-frequency data. *J. Finan. Econom.*, **4**, 353–384.

[7] Bollerslev, T., & Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *J. Econom.*, **109**, 33–65.

[8] Broadie, M., & Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Op. Res.*, **54**, 217–231.

[9] Carr, P., Geman, H., Madan, D. B. & Yor, M. (2007). Self decomposabilty and option pricing. *Math. Finance*, **17**, 31–57.

[10] Chopin, N. (2007). Inference and model choice for sequentially ordered hidden Markov models. *J. R. Statist. Soc. B*, **69**, 269–284.

[11] Del Moral, P., Doucet, A. & Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

[12] Del Moral, P., Doucet, A. & Jasra, A. (2006). Sequential Monte Carlo for Bayesian computation (with discussion). *Bayesian Statistics 8*, Ed. Bayarri, S., Berger, J. O., Bernardo, J. M., Dawid, A. P., Heckerman, D. Smith, A. F. M. and West, M., Oxford:OUP, 115–148.

[13] Del Moral, P., Doucet, A. & Jasra, A. (2008). On adaptive Sequential Monte Carlo methods. Technical Report, INRIA.

[14] Doucet, A., De Freitas, J. F. G. & Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer: New York.

[15] Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, **68**, 1343–1376.

[16] Eraker, B., Johannes, M., & Polson, N. (2003). The impact of jumps in volatilities and returns. *J. Finance*, **58**, 1269–1300.

[17] Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *J. Finance*, **59**, 1367–1403.

[18] Frühwirth, S., & Sögner, L. (2007). Bayesian estimation of the Heston stochastic volatility model. Technical Report, Johannes Kepler University.

[19] Gander, M. P. S. & Stephens, D. A. (2007). Stochastic volatility modelling with general marginal distributions: Inference, prediction and model selection. *J. Statist. Plan. Infer.*, **137**, 3068–3081.

[20] Gander, M. P. S. & Stephens, D. A. (2007). Simulation and inference for stochastic volatility models driven by Lévy processes. *Biometrika*, **94**, 627–646.

[21] Griffin, J., & Steel, M. (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *J. Econom.*, **134**, 605–644.

[22] Heston, S. L. (1993). A closed form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Fin.*, **6**, 327–334.

[23] Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance.*, **41**, 281–300.

[24] Jasra, A., Stephens, D. A. & Holmes, C. C. (2007). On population-based simulation. *Statist. Comp.*, **17**, 263–279.

[25] Kalogeropoulos, K., Dellaportas, P. & Roberts, G. O. (2007). Likelihood-based inference for correlated diffusions. Technical Report, University of Cambridge.

[26] Li, H., Wells, M. T. & Yu, C. (2008). A Bayesian analysis of return dynamics with Lévy jumps, *Rev. Fin.* (to appear).

[27] Lin, L., Liu, K. F. & Sloan, J. (2000). A noisy Monte Carlo algorithm. *Phys. Rev. D*, **61**, 074505.

[28] Madan, D. B., Carr, P., & Chang, E. (1998). The variance gamma process and option pricing. *Eur Finance Rev.*, **2**, 79–105.

[29] Mijatović, A., & Albanese, C. (2008). A stochastic volatility model for risk reversals in foreign exchange. *Int J. Theor. Appl. Fin*, (to appear).

[30] Papaspiliopoulos, O., Roberts, G. O., & Skold, M. (2003). Non-centred parameterisations for hierarchical models and data augmentation (with discussion). *Bayesian Statistics 7*, Ed. Bayarri, S., Berger, J. O., Bernardo, J. M., Dawid, A. P., Heckerman, D. Smith, A. F. M. and West, M.

[31] Pitman, J. & Yor, M. (1982). A decomposition of Bessel bridges. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **59**, 425-457.

[32] Roberts, G. O. & Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603-621.

[33] 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.

[34] Yu, J., (2005). On leverage in a stochastic volatility model. *J. Econom.*, **127**, 165178.