Bayesian Analysis (2011) **6**, Number 4, pp. 901–926

Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility

Eleni-Ioanna Delatola* and Jim E. Griffin†

Abstract. This paper presents a method for Bayesian nonparametric analysis of the return distribution in a stochastic volatility model. The distribution of the logarithm of the squared return is flexibly modelled using an infinite mixture of Normal distributions. This allows efficient Markov chain Monte Carlo methods to be developed. Links between the return distribution and the distribution of the logarithm of the squared returns are discussed. The method is applied to simulated data, one asset return series and one stock index return series. We find that estimates of volatility using the model can differ dramatically from those using a Normal return distribution if there is evidence of a heavy-tailed return distribution.

Keywords: Dirichlet process, Asset Return, Stock Index, Off-set mixture representation, Mixture model, Centred representation

1 Introduction

The prices of financial assets are usually thought to behave according to some stylized facts. The clearest example is that returns (the changes in the log price over a specified period) have a standard deviation (or volatility) that changes over time. The main purpose of stochastic volatility (SV) models is to describe this underlying time-varying volatility. The first publication of a direct time-varying volatility model was by Taylor (1982) who modelled the log of volatility by a latent AR(1) process. In line with the work of Clark (1973) and Tauchen and Pitts (1983), the use of the AR process can be explained as the representation of a random and uneven flow of new information, which is hard to be modelled in a direct way. Taylor's model for returns $y_1, y_2, \ldots, y_n$ is

$$y_t = \beta e^{h_t/2}\epsilon_t \tag{1}$$

$$h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \sigma_\eta \eta_t$$

*School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK, ed62@kent.ac.uk

†School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK, J.E.Griffin-28@kent.ac.uk

 DOI:10.1214/11-BA632

where h_t is the log-volatility at time t, and ϵ_t and η_t are independent with mean zero and unit variance. The log volatility process is an AR(1) process with persistence parameter ϕ (which must be between -1 and 1 for the process to be stationary), which implies that the stationary distribution of h_t has mean μ and variance $\sigma_\eta^2/(1-\phi^2)$. The model is unidentified but setting $\beta = 1$ or $\mu = 0$ leads to an identified model. In this case, β or $\exp(\mu/2)$ can be interpreted as the modal instantaneous volatility.

Commonly, ϵ_t and η_t are assumed Normally distributed. However, the assumption of Normality for ϵ_t has been questioned in the literature and heavier tailed distributions have been proposed as alternatives. For example, Nakajima and Omori (2009) and Jacquier et al. (2004) consider the t-distribution, Barndorff-Nielsen (1997) uses the Normal-Inverse Gaussian, Mahieu and Schotman (1998) use a mixture of Normals and Abanto-Valle et al. (2010) apply scale mixture of Normals using different mixing parameters.

A well-known problem with Bayesian inference in the SV model in equation (1) is that the likelihood has an intractable form. There are many computational schemes in the literature that deal with this problem by including the log-volatilities $h_1, h_2, \ldots, h_n$ in a Markov chain Monte Carlo (MCMC) sampler (see Broto and Ruiz (2004) for a review). Jacquier et al. (1994) proposed a single state cyclic Metropolis-Hastings algorithm to update the log-volatilities one at a time. Shephard and Kim (1994) showed that when the persistence ϕ takes values close to unity and the variance σ_η^2 takes very small values then the log-volatilities are highly correlated and the single-state algorithm generates highly correlated draws, leading to slow mixing of the sampler. Kim et al. (1998) introduced a sampling scheme which updates the log-volatilities simultaneously using a linearization of the model. Alternatively, random-length blocks of the log-volatilities in the model in (1) can be jointly updated to reduce correlation in draws (see Jensen and Maheu 2010; Abanto-Valle et al. 2010, in the context of mixture models).

The approach of Kim et al. (1998) expresses the basic SV model in equation (1) as a linear state space model by taking the logarithm of the squares of the observations, $y_t^\star = \log y_t^2$:

$$y_t^\star = h_t + z_t \qquad \text{for } t = 1, ..., n \tag{2}$$

where $z_t = \log \epsilon_t^2$. The distribution of z_t is a $\log \chi_1^2$ if ϵ_t^2 is Normally distributed and so Kalman filtering techniques cannot be directly applied. Kim et al. (1998) and Omori et al. (2007) suggest using a mixture of Normals to approximate this distribution. This allows a multi-state algorithm to be defined that updates all the log-volatilities simultaneously using a filtering forward backward sampling (FFBS) algorithm (Carter and

Kohn 1994; Frühwirth-Schnatter 1994; Durbin and Koopman 2002) leading to faster mixing than single-state algorithms. A problem with this parametrization is that the returns can take values very close to zero or even in some cases zero, leading the transformed values to be largely negative or undefined respectively. This problem is an inlier problem which can be solved by introducing an offset parameter c and defining $y_t^\star = \log(y_t^2 + c)$ (Fuller 1996).

Our contribution is to propose a Bayesian nonparametric approach to estimating the distribution of $\log\epsilon_t^2$ and so avoid parametric assumptions about the return distribution. Our model replaces the finite Normal mixture approximating the $\log\chi_1^2$ distribution with a Dirichlet process mixture model (DPM). This allows us to use the FFBS algorithm, and thus to extend the work of Kim et al. (1998) and Omori et al. (2007). We use the alternative representation of the DPM as introduced by Griffin (2010). Models with nonparametric return distributions can capture features of the return data which the parametric equivalents cannot fully capture (Gallant et al. 1997; Mahieu and Schotman 1998; Durham 2006). Durham (2006) states that, "... I find no evidence that even the simple-factor models are unable to capture the dynamics of the volatility process. The more critical problem is to capture the shape of the conditional returns distribution". From the above, one can conclude that simple parametric models will often be unable to adequately model the conditional return distribution. A DPM model offers a flexible alternative as it can be considered an infinite mixture model where the data specify the number of mixing components. Additionally, efficient algorithms have been developed that facilitate the sampling from the posterior of a DPM model (Neal 2000).

The idea of using Bayesian nonparametric methods in stochastic volatility literature has been seen in Jensen (2004) and Jensen and Maheu (2010). Jensen (2004) used a DPM prior to model the distribution of the wavelet coefficients of z_t for a fractionally integrated SV model. Jensen and Maheu (2010) introduce a semiparametric SV model which directly models the asset returns, without using the offset mixture model representation. The model we propose is a semiparametric offset mixture model, where the error terms of the observation equation are modelled nonparametrically with the DPM model and having an AR(1) process for the latent log-volatilities. In the work of Jensen and Maheu (2010), their semiparametric SV model uses the DPM representation which mixes over both the mean and the variance.

The outline of this paper is: Section 2 describes the Dirichlet process and its use in mixture modelling, Section 3 introduces the semiparametric SV model, Section 4 describes an MCMC sampler for inference, Section 5 gives a simulated example and

some empirical examples, Section 6 is a discussion.

2 Dirichlet Process Mixture model

The Dirichlet process mixture of normals model (DPM) was introduced by Lo (1984) and Ferguson (1983) and has become one of the most popular models in Bayesian nonparametrics (for other models in the Bayesian nonparametrics see Müller and Quintana (2004)). The model can be constructed in the following way. Let

$$f_k(z) = \sum_{j=1}^{k} p_j \mathrm{N}\left(z \left| \mu_j^{'}, {\sigma_j'}^2 \right.\right)$$

where $\mathrm{N}(z|\mu, \sigma^2)$ represents a Normal distribution with mean μ and variance σ^2, and $p_1, p_2, \ldots, p_k$ follow a Dirichlet distribution with parameters $\left(\frac{M}{k}, \frac{M}{k}, \ldots, \frac{M}{k}\right)$. This is a k-component normal mixture model where the location parameter $\mu_j^{'}$ and the scale parameter σ_j^2 differ from component-to-component and p_j are the mixing weights. If we further assume that $(\mu_j^{'}, {\sigma_j'}^2)$ are *a priori* independent and identically distributed, it is straightforward to show that the limit of f_k as $k \to \infty$ is well-defined and it is a Dirichlet process mixture model. Therefore, the DPM belongs in the class of infinite mixture models. Alternatively, the k-component normal mixture model can be written

$$f_k(z) = \int \mathrm{N}\left(z \left| \mu^{'}, {\sigma'}^2 \right.\right) \, dG_k(\mu', {\sigma'}^2)$$

$$G_k = \sum_{j=1}^{k} p_j \delta\left(\mu_j', {\sigma_j'}^2\right)$$

where $\delta\left(x_1, x_2\right)$ represents the Dirac measure that places mass 1 on (x_1, x_2). If $\mu', {\sigma'}^2 \overset{i.i.d.}{\sim} H$, the measure G_k limits to a measure G with a Dirichlet process prior with mass parameter M and centring distribution H if $k \to \infty$. We will write this as $G \sim \mathrm{DP}(M, H)$.

Many authors have followed Escobar and West (1995) by specifying priors on $\mu_j^{'}$ and ${\sigma_j'}^2$. This must be informative for effective inference. We will use an alternative form of prior suggested by Griffin (2010) which allows non-informative prior distributions for parameters representing overall location and overall scale. The model for observation

$z_1, z_2, \ldots, z_n$ can be written

$$z_t|\mu_t' \sim \mathrm{N}\left(\mu_t^{'}, \alpha\sigma_z^2\right), \qquad t = 1, 2, \ldots, n \tag{3}$$

$$\mu_t^{'} \sim G, \qquad t = 1, 2, \ldots, n$$

$$G \sim \mathrm{DP}\left(M, H\right)$$

$$H \equiv \mathrm{N}\left(\mu_0, \left(1-\alpha\right)\sigma_z^2\right), \tag{4}$$

where μ_0 is the overall location, σ_z^2 is the overall scale and α is interpreted as a smoothness parameter. The data, z_t, are conditionally Normally distributed and only the means $\mu_j^{'}$ are distributed as a Dirichlet process with mass parameter M. This model assumes that the mixing component variances are constant where σ_z^2 is the prior variance of z_t. The prior expected distribution of z_t is Normal with mean μ_0 and variance σ_z^2 which centres the prior of the distribution of z_t.

Many efficient MCMC sampling schemes for these models have been developed (Escobar and West 1995; Neal 2000). The model in (3) is a conjugate Dirichlet process and so standard MCMC methods as described by Escobar and West (1995) can be used directly. Griffin (2010) gives full details of the MCMC sampler needed for this model.

3 Semiparametric SV model

In this section, we present a semiparametric SV model (SPM) which has a nonparametric specification for the conditional return distribution but retains the parametric specification of the volatility process. In contrast to Jensen and Maheu (2010), the model uses the linearized parametrization of the basic SV model and so makes use of the sampling method of Kim et al. (1998) to efficiently update the volatilities in the MCMC algorithm. The SPM model is

$$y_t^\star = h_t + z_t, \tag{5}$$

$$h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \sigma_\eta\eta_t,$$

where $y_t^\star = \log\left(y_t^2 + c\right)$, $z_t = \log\epsilon_t^2 \sim F$ and $\eta_t \sim \mathrm{N}(0,1)$. By modelling the distribution of $y_t^\star$, rather than y_t directly, information about the sign of the return y_t is lost. Therefore, the distribution of y_t or ϵ_t can only be recovered by making an assumption about the distribution of the sign of the return. We will assume that the distribution of ϵ_t, and so y_t, is symmetric. Although a strong assumption, Jensen and Maheu (2010) find little evidence of skewness in their empirical examples and also find that a symmetric model gives better out-of-sample prediction performance than a model which allows skewness.

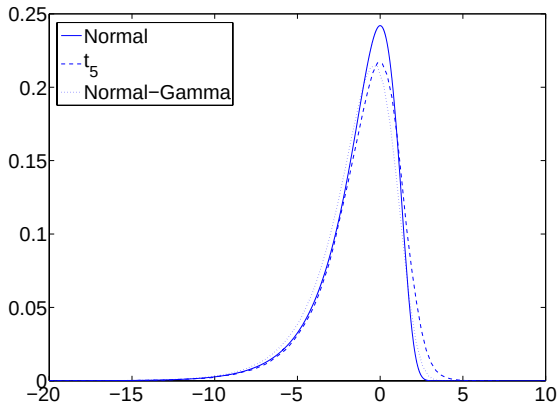

Figure 1: Plot of $\log\epsilon_t^2$ for three different distributions of ϵ_t.

To understand the types of distribution that z_t may follow, Figure 1 shows the distribution of z_t for several popular choices for ϵ_t. These are: the Normal distribution, the t distribution with 5 degrees of freedom and the Normal-Gamma distribution. The t_5 distribution has a similar left tail to the Normal distribution but it has a heavy right-hand tail. On the other hand, the Normal-Gamma distribution does not have as heavy tails as the t_5 and the distribution has a flatter left-hand tail and similar right-hand tail to the Normal distribution.

It is very common to come across zero returns when fitting these models to observed data. In the literature, different techniques have been proposed to treat these values (Harvey 1990; Sandmann and Koopman 1998). We will model the zero returns by a Normal distribution centred around $\log c$ (since $y_t^\star = \log c$ if $y_t = 0$) and model the non-zero returns by a Dirichlet process mixture of Normals. This leads to the following model for z_t,

$$p(z_t) = W\mathrm{N}\left(\log c, \sigma_0^2\right) + (1-W)\sum_{j=1}^{\infty} w_j \mathrm{N}\left(\mu_j^{'}, \alpha\sigma_z^2\right) \tag{6}$$

where W is the probability that a return is a zero return and $\sigma_0^2 = 1$ throughout the paper. As well as affecting the fit of the model, the zero returns can also lead to slow mixing of the MCMC sampling scheme if $W = 0$ (*i.e.* there is no component for the zero returns). This is because the mixture model will capture the zero returns using a component whose variance $\alpha\sigma_z^2$ will becomes very small which, in turn, leads to a large number of components. The components in the Dirichlet process mixture model

for the non-zero returns have the same variances. This is unusual for DPMs where component variances are allowed to differ. Griffin (2010) shows that the assumption of the same variances does not greatly affect inference about the unknown distribution in many circumstances and can often lead to better out-of-sample predictive performance.

If a parametric distribution is assumed for z_t, the mean of z_t is fixed and so μ is identified. However, the mean is random if z_t is given a nonparametric prior and so μ cannot be easily separated from the mean of z_t without making additional assumptions. Thus, in the MCMC sampling it is easier to subsume μ into z_t leading to the re-parameterized model

$$y_t^\star = h_t^\star + z_t^\star, \tag{7}$$

$$h_{t+1}^\star = \phi h_t^\star + \sigma_\eta \eta_t \tag{8}$$

where $h_t^\star = h_t - \mu$, $z_t^\star = z_t + \mu$ and $h_1^\star \sim \mathrm{N}\left(0, \sigma_\eta^2/(1-\phi^2)\right)$. After imposing this constraint, the mean of $h_t^\star$ is fixed to be zero and inference can be made for the SPM. This constraint has also been used by Bush and MacEachern (1996) and Jensen and Maheu (2010). It is still useful to have an estimate of μ in order to be able to compare the results of the SPM to results from the SV model in (1) with ϵ_t having a Normal distribution. This is achieved by noting that $\mathrm{E}[z_t^\star] = \mu + \mathrm{E}[z_t]$. If we assume that $\mathrm{E}[z_t] = -1.2704$ (which is chosen to match the value if ϵ_t follows a Normal distribution) which leads to the approximate value $\mu = \mathrm{E}[z_t^\star] + 1.2704$.

Our posterior inference is about the distribution of $z_t^\star$ but we are typically interested in the distribution of ϵ_t. If we assume that ϵ_t is symmetric, then the distribution of ϵ_t can be recovered from the distribution of $z_t^\star$. The variance and kurtosis of ϵ_t are also useful summaries of the distribution. The log of the variance of ϵ_t can be approximated using the expected value of $z_t^\star$ and the kurtosis of ϵ_t can be approximated using the variance of $z_t^\star$. These statements follow from taking a second order Taylor series expansion around $\mathrm{E}[\epsilon_t^2]$ and approximating the first term moments of $z_t^\star$ by

$$\mathrm{E}\left(z_t^\star\right) = \mu + \mathrm{E}\left(z_t\right) \approx \mu + \log\mathrm{E}\left(\epsilon_t^2\right) - \frac{1}{2}\frac{\mathrm{V}[\epsilon_t^2]}{(\mathrm{E}[\epsilon_t^2])^2}$$

and

$$\mathrm{V}\left(z_t^\star\right) = \mathrm{V}\left(z_t\right) \approx \frac{\mathrm{V}\left(\epsilon_t^2\right)}{\left(\mathrm{V}\left(\epsilon_t\right)\right)^2} = \mathrm{K}(\epsilon_t) + 1$$

where $\mathrm{K}(\epsilon_t)$ is the kurtosis (the fourth central moment) of ϵ_t. Therefore,

$$\mathrm{K}(\epsilon_t) \approx \mathrm{V}\left(z_t^\star\right) - 1. \tag{9}$$

and

$$\exp\{\mu\}\mathrm{V}\left(\epsilon_t\right) \approx \exp\{\mathrm{E}\left(z_t^\star\right) + \mathrm{V}\left(z_t^\star\right)/2\}.$$

The formula for μ and the expression for $\mathrm{V}\left(\epsilon_t\right)$ and $\mathrm{K}\left(\epsilon_t\right)$ all involve $\mathrm{E}[z_t^\star]$ and $\mathrm{V}[z_t^\star]$. These are hard to calculate from an MCMC sampler since they involve infinite sums. We will use the following approximations:

$$\mathrm{E}[z_t^\star|\psi] \approx W\mathrm{log}c + (1-W)\left(\sum_{j=1}^{k}\frac{n_j}{n+M}\mu_j^{'} + \frac{M}{M+n}\mu_0\right)$$

and

$$\mathrm{V}[z_t^\star|\psi] = W\left((\mathrm{log}c)^2+1\right) + (1-W)\frac{\sum_{j=1}^{k}\left(n_j\left(\alpha\sigma_z^2+\mu_j^{'2}\right)\right) + M\left(\mu_0^2+\sigma_z^2\right)}{n_2^{'}+M} \tag{10}$$

$$-\left(\mu - 1.2704\right)^2. \tag{11}$$

where $\phi = (W, c, k, n_1, \ldots, n_k, \mu_1', \ldots, \mu_k', \mu_0, a, \sigma_z^2, M)$ and n_j is the number of observations allocated to the j^{th} non-zero cluster and k is the number of distinct non-zero clusters. The posterior expectation, $\mathrm{E}[z_t^\star|y]$, can be calculated using the estimator $\widehat{\mathrm{E}}[z_t^\star|y] = \frac{1}{N}\sum_{j=1}^{N}\mathrm{E}\left[z_t^\star|\psi^{(j)}\right]$ where $\psi^{(1)}, \ldots, \psi^{(N)}$ is an MCMC sample from the posterior distribution. Similarly $\mathrm{V}\left[z_t^\star|y\right]$ can be estimated by $\frac{1}{N}\sum_{j=1}^{N}\mathrm{V}\left[z_t^\star|\psi^{(j)}\right] + \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathrm{E}\left[z_t^\star|\psi^{(j)}\right] - \widehat{\mathrm{E}}[z_t^\star|y]\right)^2$.

We assume the following priors for the parameters of the SPM model:

$$\phi \sim \mathrm{N}\left(0, 10\right) \times I_{[-1,1]}, \qquad \sigma_\eta^2 \sim \mathrm{IG}\left(2.5, 0.025\right),$$

where $\mathrm{IG}\left(a, b\right)$ is an inverted Gamma distribution with mean (if $a > 1$) $\frac{b}{a-1}$ and variance (if $a > 2$) $\frac{b^2}{(a-1)(a-2)}$ (as in Kim et al. (1998)) and $\mathrm{N}\left(\mu, \sigma^2\right) \times I_{[a,b]}$ represents a Normal distribution with mean μ and variance σ^2 restricted to the interval (a, b) and so imposes stationarity of the log-volatility process (as in Jacquier et al. (2004)). The prior for W follows a Beta distribution, $W \sim \mathrm{Be}\left(0.1, 0.9\right)$. The mass parameter of the Dirichlet process, M, has the prior suggested by Griffin and Steel (2004),

$$p\left(M\right) = \theta^\lambda \frac{\Gamma\left(2\lambda\right)}{\left(\Gamma\left(\lambda\right)\right)^2}\frac{M^{\lambda-1}}{\left(M+\theta\right)^{2\lambda}}$$

where θ denotes a prior sample size and λ is a variance parameter. While Griffin (2010) suggests a sampler for updating α, in this paper we keep the value fixed using two different values for $\alpha = 0.01$ and $\alpha = 0.05$ as suggested by Griffin (2010), to show its effect in the density estimation process. The effect of the different priors for the DPM is discussed in Griffin (2010).

4 MCMC algorithm

In this section we briefly discuss the steps of the MCMC algorithm to fit the SPM model. More details of the algorithm can be found in Appendix A. Let, $\mathbf{y}^\star = (y_1^\star, y_2^\star, ..., y_n^\star)$, $\mathbf{h} = (h_1, h_2, ..., h_n)$ and $\boldsymbol{\mu}' = \left(\mu_1', \mu_2', ..., \mu_n'\right)$. Indicator variables $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ are introduced to allocate the observations to the distinct values of the Dirichlet process. We propose the following steps for the algorithm to update the parameters:

- Initialize $\phi, \sigma_\eta^2, \sigma_z^2, \mu_0, \boldsymbol{\mu}', M, \mathbf{s}$, and W.
- Sample $\mathbf{h}|\mathbf{y}^\star, \phi, \sigma_\eta^2, \sigma_z^2, \mu_0, \boldsymbol{\mu}', M, W, \mathbf{s}$.
- Sample $\mathbf{s}|\mathbf{y}^\star, \phi, \sigma_\eta^2, \sigma_z^2, \mu_0, \boldsymbol{\mu}', M, W, \mathbf{h}$.
- Sample $\sigma_z^2, \mu_0, \boldsymbol{\mu}', M, W|\mathbf{y}^\star, \phi, \sigma_\eta^2, \mathbf{s}, \mathbf{h}$.
- Sample $\phi, \sigma_\eta^2|\mathbf{y}^\star, \sigma_z^2, \mu_0, \boldsymbol{\mu}', M, W, \mathbf{s}, \mathbf{h}$.

The full algorithm is described in Appendix A but the steps are briefly described here.

Updating Log-Volatilities

The representation of the SV model in (5) and the choice of a mixture of Normals for z_t enables us to make easier inference. Conditional on $\mathbf{s}$, the model for $\mathbf{y}^\star$ is a Gaussian dynamic linear model and so the log-volatilities $\mathbf{h}$ can be updated simultaneously using the forward filtering backward sampling (FFBS) algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994; Durbin and Koopman 2002).

Updating Mixture Components

The allocation variables $\mathbf{s}$ and the parameters of the mixture model can be updated using the methods described by Griffin (2010) which uses standard methods for Dirichlet process mixture models.

Updating SV parameters

We have found that it is necessary to mix between two parameterizations of the model when updating the SV parameters. This is because the log-volatilities $\mathbf{h}$ and the parameters $\left(\phi, \sigma_\eta^2, \mu\right)$ are usually highly correlated. Two main approaches for improving

the mixing have been proposed in the literature: reparametrize the model into a form that reduces the dependence between the parameters and the log-volatilities (Gelfand et al. 1995; Papaspiliopoulos et al. 2007) or to use a scheme that updates jointly the parameters and the log-volatilities (Kim et al. 1998; Chib et al. 2002; Andrieu and Roberts 2009). In this paper, we will use the first solution. In particular, we introduce a centred and non-centred parameterization (Gelfand et al. 1995) of the SPM model and update the parameters using a hybrid Gibbs sampler that chooses a parameterization with some probability at each step of the Gibbs sampler. The model in equation (8) is represented using the non-centred parameterization for μ, *i.e.* μ and the log-volatilities are a priori independent. The alternative parametrization is the centered defined by the reparameterization from $(\mu', \mathbf{h})$ to $(\mu^\star, \mathbf{h}^\star)$ where $\mu_j^\star = \mu_j' + \mu_0$ and $h_t^\star = h_t - \mu_0$.

5 Results

In this section, the semiparametric SV model (SPM) is fitted to simulated data, asset return data and stock index data. The results for the SPM model with $\alpha = 0.01$ and $\alpha = 0.05$ are compared to the results with the parametric SV model (PM) where ϵ_t follows a Normal distribution using the algorithm of Kim et al. (1998). The fit of the models will be assessed using log predictive scores as proposed by Kim et al. (1998). The average log predictive score for one-step ahead predictions is given by

$$\text{LPS} = -\frac{1}{T}\sum_{i=1}^{T} \log p\left(y_i^\star \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right)$$

where $\mathbf{y}_{1:t}^\star = (y_1^\star, y_2^\star, ..., y_t^\star)$ and $\hat{\theta}$ is an estimate of the model parameters and the one-step ahead predictive density is given by

$$p\left(y_i^\star \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right) = \int\int p\left(y_i^\star \left| h_i, \hat{\theta}\right.\right) p\left(h_i \left| h_{i-1}, \hat{\theta}\right.\right) p\left(h_{i-1} \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right) dh_i dh_{i-1}.$$

Smaller values of the LPS indicate a better fitting model. Monte Carlo approximations to the integral in $p\left(y_i^\star \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right)$ for all i can be efficiently calculated using draws from $p\left(h_{i-1} \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right)$ sampled using standard sequential Monte Carlo methods. The model parameters are $(\phi, \sigma_\eta^2, \mu)$ in the parametric model and (ϕ, σ_η^2, F) in the nonparametric model which are all estimated by their posterior means. In the sequential Monte Carlo sampler, it is useful to represent the posterior mean of F by a mixture of normals. The approximating mixture of normals is calculated by minimizing the Kullback-Leibler divergence between the approximation and the posterior mean of F.

We often find that the different models have similar LPS values. Consequently, it is useful to have a measure that concentrates on the prediction of tail events. We propose the average log predictive tail score for one-step ahead predictions which restricts attention to those events in the upper $100\alpha\%$ of the empirical distribution of the squared returns and is defined by

$$\text{LPTS}_\alpha = -\frac{1}{\sum_{i=1}^{T} \text{I}(y_i^\star > z_\alpha)} \sum_{i=1}^{T} \text{I}(y_i^\star > z_\alpha) \log p\left(y_i^\star \left| \mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right)$$

where z_α represents the upper $100\alpha\%$ point of the empirical distribution of the returns. The LPTS cannot be formally used for model selection as it is not considered a proper scoring rule (see Gneiting and Raftery 2007) but can be useful for understanding how the model performs for tail events.

The approximation in (9) implies that the kurtosis of the distribution of ϵ_t can be approximated by the variance of z_t. Therefore, the posterior distribution of $\sigma^2 = \text{V}[z_t^\star|\psi]$, given in (11), is reported in the results.

	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
Simulated	32930	32210	25738
Microsoft	55734	55349	34413
S&P 500	64274	67328	42530

Table 1: CPU times (in seconds) for the Semiparametric model with $\alpha = 0.01$ and $\alpha = 0.05$ for $c = 10^{-4}$.

CPU times in seconds of running each model and dataset for 50 000 iterations using code written in Matlab with a 2GHz Intel Core 2 Duo processor are given in Table 1. All results are calculated by running the sampler, coded in Matlab, for 200 000 iterations using two quad core Xeon 2.53Ghz CPUs. The first 100 000 draws are discarded as burn-in period and, after this period, we retain every 10^{th} draw to reduce autocorrelation between draws. The datasets that we will be using for our analysis are: simulated returns, Microsoft returns and S&P 500 returns.

5.1 Simulated Data

We generated 3000 data points from the basic SV model in (1), with $\phi = 0.97$ $\mu = 0$, $\sigma_\eta^2 = 0.0225$ and ϵ_t following a Student t-distribution with 7 degrees of freedom. The data used are shown in Figure 2.

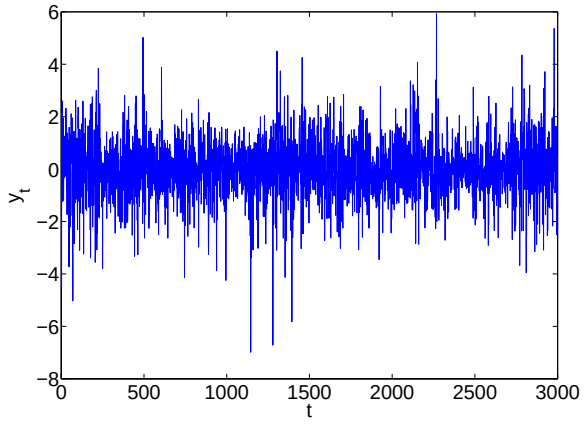

Figure 2: Simulated Data: the actual values of y_t.

Figure 3 presents the autocorrelation function for sampled values from each parameter of SPM with $\alpha = 0.01$. The autocorrelations decay quickly for all parameters apart from M and σ^2 but these parameters do not show a large degree of autocorrelation. This suggests that the SPM sampler mixes quickly.

	True	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
ϕ	0.97	0.960 *(0.928, 0.979)*	0.965 *(0.938, 0.982)*	0.904 *(0.841, 0.942)*
μ	0.000	0.023 *(-0.143,0.197)*	0.095 *(-0.07,0.269)*	0.267 *(0.135,0.403)*
σ_η	0.15	0.161 *(0.116,0.232)*	0.145 *(0.100,0.202)*	0.304 *(0.230,0.404)*
σ^2		5.482 *(5.009,6.115)*	4.929 *(4.778,5.082)*	
M		0.360 *(0.070,1.144)*	2.622 *(0.445,17.453)*	
k		4 *(2,7)*	19 *(5,90)*	
W		0.028 *(0.012,0.043)*	0.000 *(0.000,0.003)*	

Table 2: Simulated Data: Posterior medians and 95% Credible Intervals for the Semi-parametric (SPM) and Parametric (PM) for $c = 10^{-4}$.

Table 2 contains the posterior estimates for the parameters of PM and SPM with $\alpha = 0.01$ and $\alpha = 0.05$. PM poorly estimates all parameters of the model with ϕ underestimated, and μ and σ_η overestimated. In contrast, both instances of the semi-parametric model estimate all parameters of the model well. The t_7 distribution has a kurtosis of 5. The posterior median kurtosis is approximately 4.5 when $\alpha = 0.01$ and 3.9 when $\alpha = 0.05$. Therefore, the kurtosis is a little underestimated but still has a

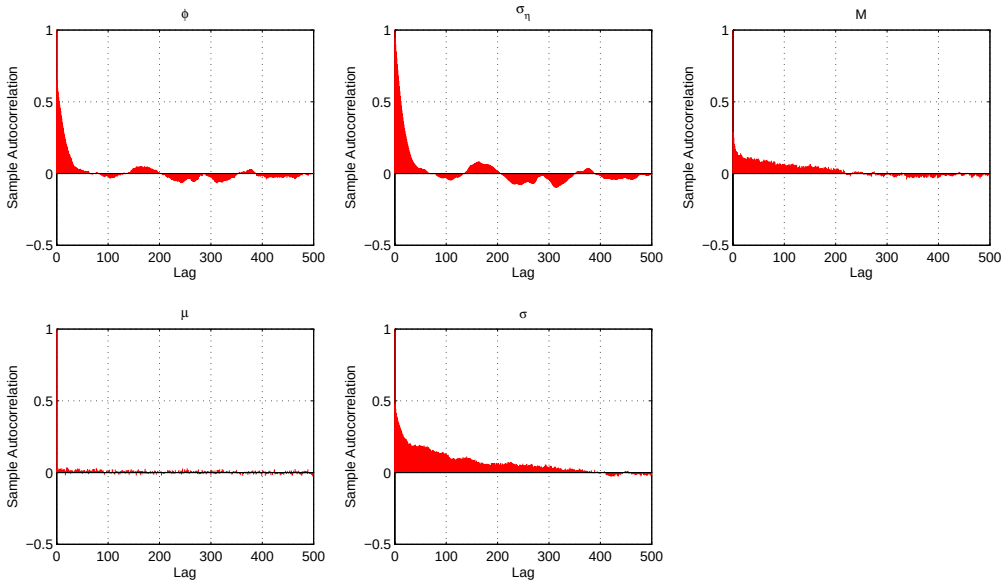

Figure 3: Autocorrelation Plot for SPM with $\alpha = 0.01$

value that is larger than the value associated with a Normal distribution. We also used the method of Nakajima and Omori (2009) to fit a model with a t-distribution (with unknown degrees of freedom). The results are very similar to those for the SPMs with the posterior median kurtosis being 4.5.

Figure 4 shows the fitted volatilities for both the SPM and the PM for $c = 10^{-4}$ and the posterior mean of the density of $\log\epsilon_t^2$. Panel (a) in Figure 4 shows that the posterior means of the volatilities under PM tend to undersmooth the volatilities with extra peaks introduced. The SPM models are able to estimate the volatilities well. We might expect that the semi-parametric model leads to larger uncertainty in the estimates of the volatilities since the model is less restricted. Panels (c) and (d) of Figure 4 show the credibility intervals for the SPM models which are very similar to those for a model with a t-distribution for ϵ_t. Panel (E) of Figure 4 shows the predictive distribution of the SPM using different values of α. Both choices lead to a satisfactory fit to the true distribution used to generate the data. The lack of fit of the Normal distribution used in PM is shown by the sharper decline of the density in the right-hand tail.

Table 3 shows the log predictive scores for SPM with $\alpha = 0.01$ and $\alpha = 0.05$, and for

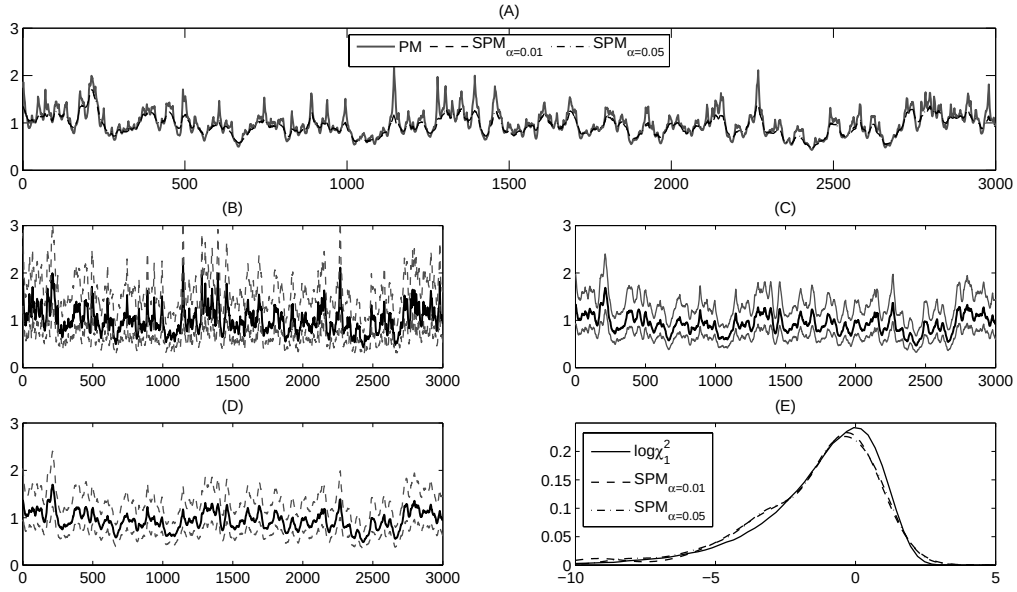

Figure 4: Simulated Data: Posterior mean volatilities for the PM and SPMs (Panel A); posterior median and 95% credible interval for: PM (Panel B), SPM ($\alpha = 0.05$) (Panel C) and SPM ($\alpha = 0.01$) (Panel D); Posterior mean distributions (Panel E).

PM. Overall, the LPS is smaller for the SPMs for all the offsets examined. However, the differences between the models tend to be small. The results for the two tail measures show that $\text{LPTS}_{0.05}$ is similar under the PM and the SPMs but the SPMs outperform the PM for $\text{LPTS}_{0.01}$. We found that small values of the offset, for example $c = 10^{-7}$, can lead to poor predictive performance for tails events and this suggests that a larger value of c would lead to more reliable results. Therefore, we use $c = 10^{-4}$ in the following examples.

5.2 Real data examples

We fitted the SPM and PM models to the compounded returns in percentages (which is $y_t = 100\log(r_t/r_{t-1})$ where r_t is the price at time t) of Microsoft (MSFT) from January 4, 1993 to December 31, 2008, which has 4030 data points, and the Standard and Poors 500 (S&P 500) index from March 13, 1980 to June 6, 2000, which has 5136 data points.

	Offset	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
LPS	$c = 10^{-3}$	2.10	2.09	2.13
	$c = 10^{-4}$	2.14	2.14	2.15
	$c = 10^{-7}$	2.15	2.15	2.16
$\text{LPTS}_{0.05}$	$c = 10^{-3}$	3.06	3.08	2.99
	$c = 10^{-4}$	3.13	3.06	3.00
	$c = 10^{-7}$	3.12	3.13	2.99
$\text{LPTS}_{0.01}$	$c = 10^{-3}$	4.36	4.50	4.60
	$c = 10^{-4}$	4.43	4.13	4.60
	$c = 10^{-7}$	4.41	4.42	4.58

Table 3: Simulated Data: Log-Predictive Scores for competing models: PM, and SPM with $\alpha = 0.01$ and $\alpha = 0.05$.

The data are depicted in Figure 5.

	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
ϕ	0.999 *(0.996,0.999)*	0.996 *(0.991,0.999)*	0.978 *(0.967,0.987)*
μ	1.291 *(-0.204, 3.311)*	1.187 *(0.442, 2.120)*	1.139 *(0.837, 1.439)*
σ_η^2	0.059 *(0.051, 0.067)*	0.078 *(0.061, 0.103)*	0.200 *(0.165, 0.240)*
σ^2	5.421 *(5.299, 5.539)*	5.546 *(5.395,5.722)*	
M	22.96 *(11.39, 44.00)*	33.69 *(9.67, 73.14)*	
k	123 *(72,197)*	141 *(55, 283)*	
W	0.017 *(0.019, 0.023)*	0.016 *(0.010, 0.022)*	

Table 4: MSFT: Posterior medians and 95% credible intervals for SPM with $\alpha = 0.01$ and $\alpha = 0.05$ and PM for $c = 10^{-4}$.

Table 4 contains the results for the MSFT returns. The posterior median of ϕ is larger under the SPMs and the posterior median of σ_η^2 is smaller under the SPMs. This suggests that the PM tends to undersmooth the estimated volatilities and over-reacts to large absolute returns (and shows a similar pattern to the results for the simulated data). Although the posterior medians of ϕ for the SPMs are close to 1, this does not imply violation of the stationarity assumption. With the SPMs, the parameter μ is smaller, suggesting a smaller variance for ϵ_t and σ^2 is estimated to be much larger than 4 suggesting that the distribution of ϵ_t is heavy tailed. This suggests that the PM again over-estimates the variance to compensate for the lighter tails of the assumed return distribution. Finally, the posterior median of the number of clusters k is relatively big

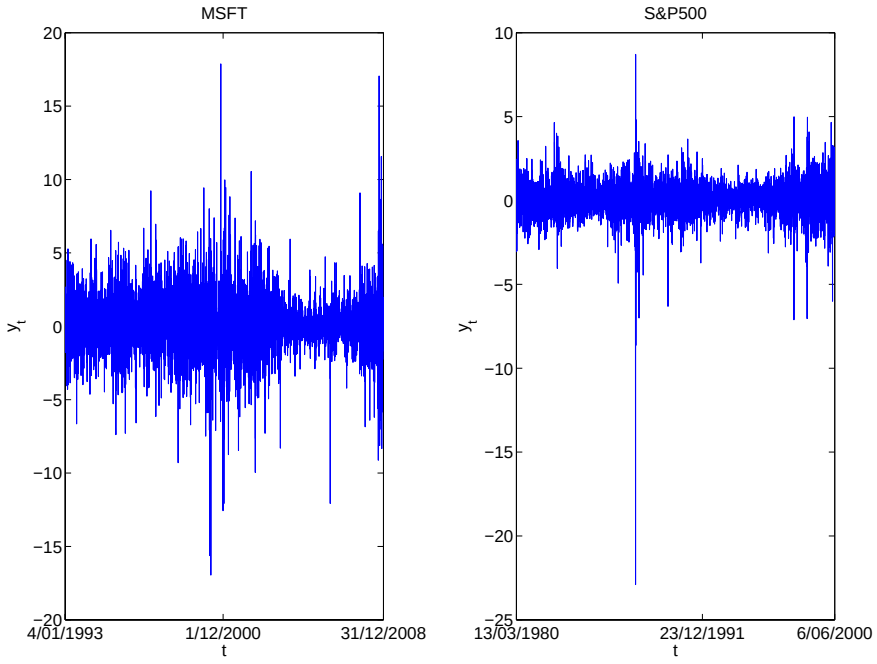

Figure 5: The returns of (a) Microsoft and (b) the S & P 500 index.

while the posterior median of W indicates the existence of zero returns. These findings are similar to those found by Jensen and Maheu (2010).

Other posterior summaries for the MSFT returns are shown in Figure 6. The posterior mean of the density of $\log\epsilon_t^2$ is shown in panel (E) and shows a bimodal distribution. The larger mode represents the main body of the data but a second mode is introduced at a much smaller value of $\log\epsilon_t^2$. This secondary mode is caused by the larger number of zero log-returns. The position of the mode is sensitive to the choice of c since a zero return is recorded as $\log c^2$. The posterior distributions of the volatilities are shown in the other panels and illustrate a smoother estimate for the SPMs compared to the PM.

The predictive performance of the SPMs and PM for the MSFT asset returns series are shown in Table 5. The results regarding the LPS score are qualitively the same for all the offset parameters apart from $c = 10^{-7}$ where the PM is favoured. That is why we employ the LPTS to see how the models behave in extreme events. The difference is much more pronounced for the tail scores which substantially favour the SPMs for

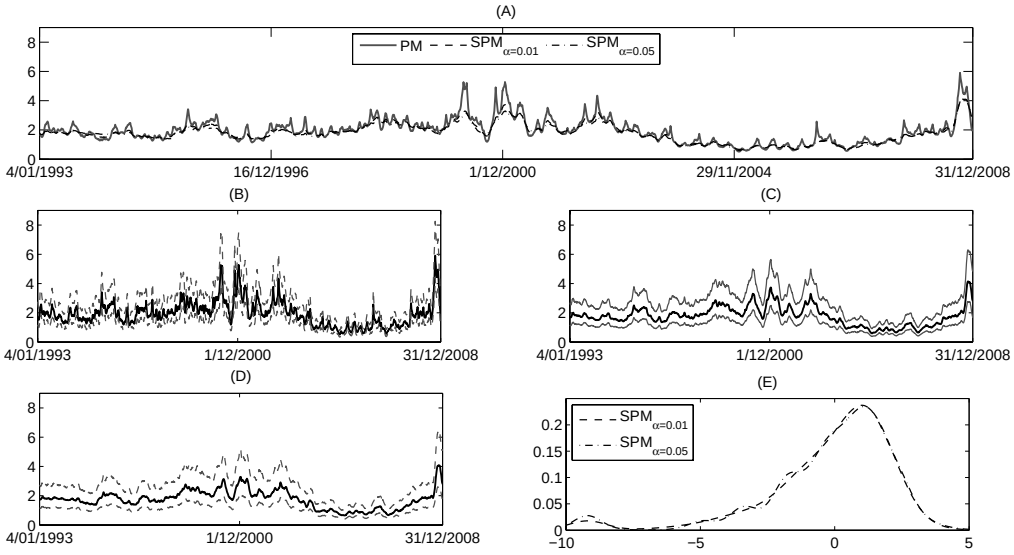

Figure 6: MSFT: Posterior mean volatilities for the PM and SPMs (Panel A); posterior median and 95% credible interval for: PM (Panel B), SPM ($\alpha = 0.05$) (Panel C) and SPM ($\alpha = 0.01$) (Panel D); Posterior mean distributions (Panel E).

$c = 10^{-3}$ and $c = 10^{-4}$. In the case of $c = 10^{-7}$, the PM seems to be doing a better job than the SPM.

Results of applying these models to a further asset price (General Motors from January 2, 1980 to December 31, 1996) are presented in Delatola and Griffin (2010) and show similar features to those identified in this analysis.

Table 6 shows the results of fitting the SPMs and the PM to the S & P 500 data. The inference has similar features to the inference for the asset return data. The posterior median of the persistence parameter, ϕ, is estimated to be larger for the SPMs and the variance of the volatility equation, σ_η^2, was estimated to be smaller (again, the 95% credible intervals for σ_η^2 with the PM and SPMs do not cross). As in the case of the Microsoft returns, the posterior median for ϕ for the SPM model might be close to 1, but this does not imply a violation of the stationarity. The estimated value of σ^2 is smaller than with the asset return data and suggests that the return distribution has a lower level of kurtosis.

	Offset	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
LPS	$c = 10^{-3}$	2.13	2.16	2.13
	$c = 10^{-4}$	2.18	2.18	2.16
	$c = 10^{-7}$	2.26	2.50	2.21
$\text{LPTS}_{0.05}$	$c = 10^{-3}$	3.99	3.88	4.62
	$c = 10^{-4}$	4.39	4.57	5.40
	$c = 10^{-7}$	8.50	18.57	7.45
$\text{LPTS}_{0.01}$	$c = 10^{-3}$	3.95	3.61	4.79
	$c = 10^{-4}$	4.14	4.09	5.90
	$c = 10^{-7}$	10.61	26.24	9.10

Table 5: MSFT: Log-Predictive Score for competing models: PM and SPM with $\alpha = 0.01$ and $\alpha = 0.05$.

	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
ϕ	0.994 *(0.989,0.998)*	0.994 *(0.989,0.998)*	0.982 *(0.972,0.990)*
μ	-0.193 *(-0.704, 0.303)*	-0.213 *(-0.639,0.251)*	-0.348 *(-0.556,-0.096)*
σ_η	0.072 *(0.057,0.090)*	0.073 *(0.058,0.091)*	0.144 *(0.113,0.173)*
σ^2	4.239 *(3.346,7.116)*	3.592 *(3.320,4.621)*	
M	0.621 *(0.135,1.771)*	2.195 *(0.447, 15.407)*	
k	6 *(3, 13)*	17 *(6, 46)*	
W	0.037 *(0.001, 0.047)*	0.039 *(0.001, 0.118)*	

Table 6: S & P 500: Posterior medians and 95% credible intervals for SPM with $\alpha = 0.01$ and $\alpha = 0.05$ and PM for $c = 10^{-4}$.

Figure 7 shows the estimated volatilities and posterior mean of the distribution of $\log\epsilon_t^2$. The posterior mean volatilities and the credibility intervals for the volatility are very similar for the PM and SPMs, indicating the similarity of the asset return distribution for the SPMs and the PM.

As far as predictability is concerned, the results in Table 7 indicate that both the PM and the SPMs seem to have the same ability for the S & P 500 data with LPS. The tail LPS scores show a slightly different picture. With the S & P 500 data, the $\text{LPTS}_{0.05}$ shows better predictive performance for the SPMs, the same can be seen for the $\text{LPTS}_{0.01}$.

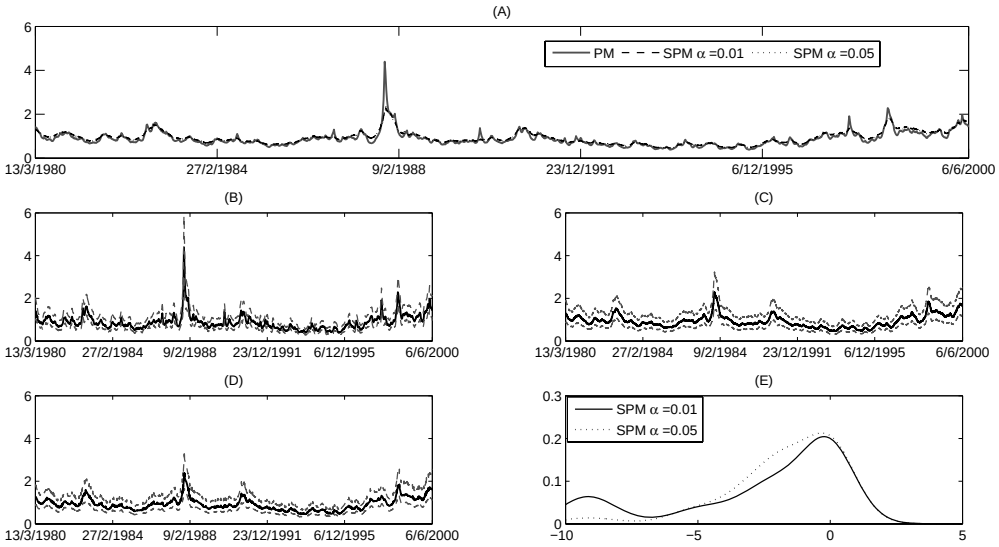

Figure 7: S & P 500: Posterior mean volatilities for the PM and SPMs (Panel A); posterior median and 95% credible interval for: PM (Panel B), SPM ($\alpha = 0.05$) (Panel C) and SPM ($\alpha = 0.01$) (Panel D); Posterior mean distributions (Panel E).

6 Discussion

This paper presents a method for Bayesian semiparametric inference in stochastic volatility models. The volatility equation is given a parametric form and the return distribution is modelled nonparametrically. The method models $\log\epsilon_t^2$ rather than ϵ_t (as considered by Jensen and Maheu (2010)). This allows efficient computational methods using forward-filtering backward-smoothing methods to be applied to the difficult problem of updating the log-volatilities in a Gibbs sampler. We discuss links between the mean and variance of $\log\epsilon_t^2$ and the variance and kurtosis of ϵ_t which allows us to simply interpret inference from our model. The results of fitting the model to data suggest that the model can give very different estimates of volatility to the standard parametric model with a Normal return distribution when there is evidence that the return distribution has heavy tails. In this case, the parametric model substantially undersmooths volatility. Out-of-sample results show that the model gives much better prediction than the parametric model in certain cases, particularly asset returns which seem to have heavier than Normal tails. We also show that the model gives similar

	Offset	SPM ($\alpha = 0.01$)	SPM ($\alpha = 0.05$)	PM
LPS	$c = 10^{-3}$	2.10	2.10	2.10
	$c = 10^{-4}$	2.24	2.13	2.12
	$c = 10^{-7}$	2.13	2.13	2.13
$\text{LPTS}_{0.05}$	$c = 10^{-3}$	3.56	3.54	3.96
	$c = 10^{-4}$	3.54	4.40	4.55
	$c = 10^{-7}$	4.70	4.70	4.85
$\text{LPTS}_{0.05}$	$c = 10^{-3}$	3.79	3.77	4.04
	$c = 10^{-4}$	3.97	4.48	5.12
	$c = 10^{-7}$	6.21	6.25	6.52

Table 7: S & P 500: Log-Predictive Scores for competing models: PM, SPM with $\alpha = 0.01$ and $\alpha = 0.05$.

results to the parametric model if the return distribution is Normal or close to Normal.

The current paper is a first step in the development of flexible semiparametric Bayesian models for stochastic volatility in discrete time which can be fitted efficiently. The model currently lacks a leverage effect and the volatility equation is given a simple form. Future work will consider the addition of a leverage effect and further nonparametric modelling of the volatility equation.

Appendix: MCMC algorithm for SPM model

In this section, we give in detail the MCMC scheme for the SPM.

- Initialize: ϕ, σ_η^2, μ_0, σ_z^2, W, $\mathbf{s}$ and $\boldsymbol{\mu}'$. We denote $\mathbf{s} = \{s_t\}_{t=1}^n$, $\mathbf{h} = \{h_t\}_{t=1}^n$ and $\boldsymbol{\mu}' = \left\{\mu_i'\right\}_{i=1}^k$ where k is the number of clusters.
- Updating $\mathbf{h}|\mathbf{y}^\star, \phi, \sigma_\eta^2, \mathbf{s}, \sigma_z^2$. The log-volatilities, $\mathbf{h}$, can be updated jointly using the FFBS algorithm presented in Kim et al. (1998).
- Updating $\mathbf{s}$: The indicator variables can be drawn using inversion sampling,

$$p\left(s_t = i\right) \propto \begin{cases} W\frac{1}{\sqrt{\sigma_0^2}}\exp\left\{-\frac{1}{2\sigma_0^2}\left(y_t^\star - h_t - \log c\right)^2\right\} & i = 1 \\ (1-W)\frac{n_i^{-t}}{\sqrt{\alpha\sigma_z^2}\left(n_2'-1+M\right)}\exp\left\{-\frac{1}{2\alpha\sigma_z^2}\left(y_t^\star - h_t - \mu_i'\right)^2\right\} & 2 \le i \le k^{-t} \\ (1-W)\frac{M}{\sqrt{\sigma_z^2}\left(n_2'-1+M\right)}\exp\left\{-\frac{1}{2\sigma_z^2}\left(y_t^\star - h_t - \mu_0\right)^2\right\} & i = k^{-t}+1. \end{cases}$$

Let $n_i^{-t} = \sum_{i=1;j\neq t}^{n} I\left(s_i = k\right)$ and k^{-t} be the number of clusters excluding the observation at time t. Also, denote by $n_2^{'}$ the number of observations that belong to the non-zero return cluster. In the above scheme, we sample the indicator variables, where $s_t = 1$ indicates that the observation belongs to the zero-return cluster and belongs to the non-zero cluster for all other values of s_t .

- Choosing parametrizations. With probability 0.5 each, choose between updating the parameters using the centered or non-centered parametrization of the SV model.

Non-centred Parametrization

– Updating $\mu_i^{'}$. For each of the nonzero components, the full conditional distribution follows a Normal distribution :

$$\mathrm{N}\left(\frac{\frac{\sum_{j|s_j=i}\left(y_j^{\star}-h_t\right)}{\alpha}+\frac{\mu_0}{1-\alpha}}{\frac{n_i}{\alpha}+\frac{1}{1-\alpha}}, \frac{\sigma_z^2}{\frac{n_i}{\alpha}+\frac{1}{1-\alpha}}\right).$$

– Updating μ_0. For the non-centered parametrization, μ_0 is Normal distributed with mean:

$$\sigma_{\mu_0}^2\left\{\frac{1-\phi^2}{\sigma_\eta^2}h_1+\frac{1-\phi}{\sigma_\eta^2}\sum_{t=1}^{n_2^{'}}\left(h_{t+1}-\phi h_t\right)\right\}$$

and

$$\sigma_{\mu_0}^2 = \frac{\sigma_\eta^2}{(n_2^{'}-1)(1-\phi)^2+(1-\phi^2)}.$$

– Updating σ_z^2: The full conditional of σ_z^2 follows an inverted-Gamma:

$$\mathrm{IG}\left(\frac{n_2^{'}+k}{2}, \frac{1}{2}\left[\frac{\sum_{i=1}^{n_2^{'}}\left(y_i^{\star}-\mu_{s_i}^{'}\right)^2}{\alpha}+\frac{\sum_{i=1}^{k}\left(\mu_i^{'}-\mu_0\right)^2}{1-\alpha}\right]\right).$$

– Updating ϕ. The scheme to update ϕ is a Metropolis-Hastings algorithm. We propose a value for ϕ, $\phi^{\star}$, from a truncated Normal distribution in the interval $[-1, 1]$,

$$\phi^{\star} \sim \mathrm{N}_{[-1,1]}\left(\frac{\sum_{t=1}^{n_2^{'}-1}\left(h_t-\mu_0\right)\left(h_{t+1}-\mu_0\right)}{\sum_{t=1}^{n_2^{'}-1}\left(h_t-\mu_0\right)}, \frac{\sigma_\eta^2}{\sum_{t=1}^{n_2^{'}-1}\left(h_t-\mu_0\right)}\right).$$

We accept the proposal with probability $\min(1, \exp\{g(\phi^\star) - g(\phi)\})$ where

$$g(\phi) = \log\pi(\phi) - \frac{(h_1 - \mu_0)^2(1-\phi^2)}{2\sigma_\eta^2} + \frac{1}{2}\log(1-\phi^2).$$

– Updating σ_η^2. The full conditional distribution for σ_η^2 is an inverted-Gamma

$$\sigma_\eta^2 | h_1^{'}..., h_n, \phi \sim \text{IG}\left(2.5 + \frac{n_2^{'}}{2}, \beta^\star\right)$$

where

$$\beta^\star = 0.025 + \frac{(h_1 - \mu_0)^2(1-\phi^2) + \sum_{t=1}^{n_2^{'}-1}(h_{t+1} - \mu_0) - \phi(h_t - \mu_0))^2}{2}.$$

Centred Parametrization

To apply the MCMC scheme for the centered parametrization, set $\mu_i^\star = \mu_i' + \mu_0$ and and $h_t^\star = h_t - \mu_0$.

– Updating μ_0. The full conditional for μ_0 in the centered parametrization, which concerns only the nonzero components, follows a Normal distribution:

$$\text{N}\left(\frac{\sum_{i=1}^k \mu_i^\star}{k}, \frac{(1-\alpha)\sigma_z^2}{n_2^{'}}\right).$$

– Updating σ_z^2: The full conditional of σ_z^2 follows an inverted-Gamma:

$$\text{IG}\left(\frac{n_2^{'} + k}{2}, \frac{1}{2}\left[\frac{\sum_{i=1}^{n_2^{'}}\left(y_i^\star - \mu_{s_i}^\star\right)^2}{\alpha} + \frac{\sum_{i=1}^k \left(\mu_i^\star\right)^2}{1-\alpha}\right]\right).$$

– Updating ϕ. The scheme to update ϕ is a Metropolis-Hastings algorithm. We propose a value for ϕ, $\phi^\star$, from a truncated Normal distribution in the interval $[-1, 1]$,

$$\phi^\star \sim \text{N}_{[-1,1]}\left(\frac{\sum_{t=1}^{n_2^{'}-1} h_t^\star h_{t+1}^\star}{\sum_{t=1}^{n_2^{'}-1} h_t^\star}, \frac{\sigma_\eta^2}{\sum_{t=1}^{n_2^{'}-1} h_t^\star}\right).$$

We accept the proposal with probability $\min(1, \exp\{g(\phi^\star) - g(\phi)\})$ where

$$g(\phi) = \log\pi(\phi) - \frac{h_1^{\star 2}(1-\phi^2)}{2\sigma_\eta^2} + \frac{1}{2}\log(1-\phi^2).$$

– Updating σ_η^2. The full conditional distribution for σ_η^2 is an inverted-Gamma

$$\sigma_\eta^2 | h_1^\star, ..., h_n^\star, \phi \sim \text{IG}\left(2.5 + \frac{n_2^{'}}{2}, \right)$$

where

$$\beta^\star = 0.025 + \frac{h_1^{\star 2}(1-\phi^2) + \sum_{t=1}^{n_2^{'}-1}(h_{t+1}^\star - \phi h_t^\star)^2}{2}.$$

We transform to $\mu_i' = \mu_i^\star - \mu_0$ and $h_t = h_t^\star + \mu_0$ in order to be consistent with the non-centered parametrization of the SPM.

- Updating M: The scheme used to update M is the ones used in Griffin (2010). It is an independence Metropolis-Hastings sampler.

- Updating W. The full conditional of W follows a Dirichlet distribution:

$$W \sim \text{Dir}\left(1 + (n - n_2^{'}), 4 + n_2^{'}\right).$$

- Updating μ: According to the DPM, μ can be estimated as

$$\mu = W\text{log}c + (1-W)\left(\sum_{j=1}^{k}\frac{n_j}{n+M}\mu_j^{'} + \frac{M}{M+n}\mu_0\right) + 1.2704.$$

- Updating σ^2: Similar to the step above we have,

$$\begin{aligned}\sigma^2 =& (1-W)\frac{\sum_{j=1}^{k}\left(n_j\left(\alpha\sigma_z^2 + \mu_j^{'2}\right)\right) + M\left(\mu_0^2 + \sigma_z^2\right)}{n_2^{'} + M} \\ &+ W\left((\text{log}c)^2 + 1\right) - \left(\mu - 1.2704\right)^2.\end{aligned}$$

References

Abanto-Valle, C., Bandyopadhyay, D., Lachos, V., and Enriquez, I. (2010). "Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions." *Computational Statistics and Data Analysis*, 54: 2883–2898. 902

Andrieu, C. and Roberts, G. O. (2009). "The Pseudo-Marginal Approach For Efficient Monte Carlo Computations." *Annals Of Statistics*, 37(2): 697–725. 910

Barndorff-Nielsen, O. E. (1997). "Normal inverse Gaussian distributions and stochastic volatility modelling." *Scandinavian Journal Of Statistics*, 24: 1–13. 902

Broto, C. and Ruiz, E. (2004). "Estimation methods for stochastic volatility models: a survey." *Journal of Economic Surveys*, 18: 613–649. 902

Bush, C. and MacEachern, S. (1996). "A semiparametric Bayesian model for randomised block designs." *Biometrika*, 83: 275–285. 907

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81: 541–553. 902, 909

Chib, S., Nadari, F., and Shephard, N. (2002). "Markov chain Monte Carlo methods for stochastic volatility models." *Journal of Econometrics*, 108: 281–316. 910

Clark, P. K. (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica*, 41: 135–55. 901

Delatola, E.-I. and Griffin, J. E. (2010). "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility." Technical Report UKC/SMSAS/10/012, School of Mathematics, Statistics and Actuarial Science, University of Kent. 917

Durbin, J. and Koopman, S. J. (2002). "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89(3): 603–616. 903, 909

Durham, G. B. (2006). "Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models." *Journal of Econometrics*, 133: 273–305. 903

Escobar, M. D. and West, M. (1995). "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association*, 90: 577–588. 904, 905

Ferguson, T. S. (1983). "Bayesian Density Estimation by Mixtures of Normal Distributions." In Rizvi, M. H., Rustagi, J., and Siegmund, D. (eds.), *Recent Advances In Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, 287–302. New York: Academic Press. 904

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15: 183–202. 903, 909

Fuller, W. A. (1996). *Introduction to statistical time series*. New York: Wiley. 903

Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). "Estimation of stochastic volatility models with diagnostics." *Journal of Econometrics*, 81: 159–192. 903

Gelfand, A. E., Sahu, S., and Carlin, B. (1995). "Efficient Parametrization for Normal Linear Mixed Effects Models." *Biometrika*, 82: 479–488. 910

Gneiting, T. and Raftery, A. E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association*, 102(477): 359–378. 911

Griffin, J. E. (2010). "Default priors for density estimation with mixture models." *Bayesian Analysis*, 5(1): 45–64. 903, 904, 905, 907, 908, 909, 923

Griffin, J. E. and Steel, M. F. J. (2004). "Semiparametric Bayesian inference for stochastic frontier models." *Journal of Econometrics*, 123: 121–152. 908

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. 906

Jacquier, E., Polson, N. G., and Rossi, P. (2004). "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors." *Journal of Econometrics*, 122: 185–212. 902, 908

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business & Economic Statistics*, 12: 371–389. 902

Jensen, M. J. (2004). "Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility Models." *Journal of Time Series Analysis*, 25(6): 895–922. 903

Jensen, M. J. and Maheu, J. M. (2010). "Bayesian semiparametric stochastic volatility modeling." *Journal of Econometrics*, 157: 306–316. 902, 903, 905, 907, 916, 919

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, 65: 361–393. 902, 903, 905, 908, 910, 920

Lo, A. Y. (1984). "On a class of Bayesian Nonparametric estimates: 1. density estimates." *Annals Of Statistics*, 12: 351–357. 904

Mahieu, R. J. and Schotman, P. C. (1998). "An empirical application of stochastic volatility models." *Journal of Applied Econometrics*, 13: 333–360. 902, 903

Müller, P. and Quintana, F. A. (2004). "Nonparametric Bayesian Data Analysis." *Statistical Science*, 19(1): 95–110. 904

Nakajima, J. and Omori, Y. (2009). "Leverage, heavy-tails and correlated jumps in stochastic volatility models." *Computational Statistics & Data Analysis*, 53: 2335–2353. 902, 913

Neal, R. M. (2000). "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics*, 9: 249–265. 903, 905

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference." *Journal of Econometrics*, 140: 425–449. 902, 903

Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2007). "A general framework for the parametrization of hierarchical models." *Statistical Science*, 22: 59–73. 910

Sandmann, G. and Koopman, S. J. (1998). "Estimation of stochastic volatility models via Monte Carlo maximum likelihood." *Journal of Econometrics*, 87(2): 271–301. 906

Shephard, N. and Kim, S. (1994). "Bayesian Analysis of Stochastic Volatility Models: Comment." *Journal of Business & Economic Statistics*, 12: 406–410. 902

Tauchen, G. E. and Pitts, M. (1983). "The Price Variability-Volume Relationship on Speculative Markets." *Econometrica*, 51: 485–505. 901

Taylor, S. J. (1982). "Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices 1961-79." In Anderson, O. D. (ed.), *Time Series Analysis: Theory and Practice 1*, 203–226. North-Holland. 901

Acknowledgments

The work in this paper was supported by EPSRC Doctoral Training Grant EP/P503388/1 and forms part of Eleni-Ioanna Delatola's PhD thesis.