A Bayesian Semiparametric Model for Volatility with a Leverage Effect

E.-I. Delatola[a], J.E. Griffin[a]

[a]*School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K.*

Abstract

A Bayesian semiparametric stochastic volatility model for financial data is developed. This nonparametrically estimates the return distribution from the data allowing for stylized facts such as heavy tails of the distribution of returns whilst also allowing for the leverage effect, which is the correlation between the returns and changes in volatility. An efficient MCMC algorithm is described for inference. The model is applied to simulated data and two real data sets. These show that choosing a parametric return distribution can have a substantial effect on estimation of the leverage effect.

1. Introduction

In the last couple of decades, stochastic volatility (SV) models have enjoyed great popularity for analyzing financial data. This is mainly attributed to the development of new, more advanced techniques in econometrics, as well as the availability of rapidly increasing computing power. The SV model as introduced by Taylor (1982) captured the heterogeneity of daily returns of sugar prices using a latent autoregressive process of order 1, AR(1), for the logged variance of a normal return distribution. However, this model was unable to capture other features of financial data such as heavy tails of the conditional distribution of returns, price jumps and the leverage effect. Black (1976) introduced the term leverage effect when observing that an increase in a stock price tends to lead to a smaller increase in its variance than a fall in the stock price of the same size. Ever since, many

*Corresponding author. Tel.: +44-2476-523369; fax: +44-2476-524532

Email addresses: `ed62@kentforlife.net` (E.-I. Delatola), `J.E.Griffin-28@kent.ac.uk` (J.E. Griffin)

Preprint submitted to Elsevier *June 22, 2012*

extensions of the SV model have been introduced that incorporated such stylized features. This paper will address the issue of building an SV model with a leverage effect and heavy tails within the Bayesian nonparametric framework.

Harvey and Shephard (1996) introduced a nonlinear SV model that could capture the leverage effect. Let P_t denote the daily price of an asset or a stock index at time t for $t = 1, \ldots, n$. The daily return of the asset or the stock index at time t is defined to be $y_t = \frac{P_t}{P_{t-1}} - 1$. The nonlinear SV model with leverage of Harvey and Shephard (1996) is represented as

$$y_t = \beta \exp\left(h_t/2\right) \epsilon_t, \tag{1}$$

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

where h_t is the log-volatility at time t and β is the modal instantaneous volatility, which for identification reasons is set to $\beta = 1$. The persistence parameter is ϕ, which is assumed to be $|\phi| \leq 1$ to ensure the stationarity of h_t, and μ is the overall mean of the log-volatility. Unlike earlier SV models, the error terms (ϵ_t, η_t) are independently and identically distributed according to a bivariate normal distribution with mean $\mathbf{0} = (0,0)^{'}$ and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix},$$

where σ_η^2 is the variance of the log-volatility and ρ is the correlation between the error terms. The addition of this parameter introduces correlation between the errors in the return distribution and the changes in the log-volatility and so allows the model to capture the leverage effect. The volatility at time $t = 1$ is drawn from the stationary distribution which is $h_1 \sim \mathrm{N}\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right)$, where $x \sim \mathrm{N}(m, \sigma^2)$ represents that x follows a normal distribution with mean m and variance σ^2. Omori et al. (2007) discussed Markov chain Monte Carlo (MCMC) methods for this model. They worked with $y^\star = \log(y_t + c)^2$ where c is small and so linearize the model of Harvey and Shephard (1996). The error terms of the return equation in this representation, $\log \epsilon_t^2$, which follow a $\log \chi_1^2$ distribution, can be accurately approximated by a 10-component mixture of normals. Nakajima and Omori (2009) extended the work of Omori et al. (2007) to incorporate jumps and heavy tails. An extension of the stochastic volatility to include leverage and heavy tails was also proposed by Jacquier et al. (2004) who make posterior inference using the non-linear representation of the model.

Several Bayesian nonparametric approaches to modelling the heavy tails in financial time series have recently been proposed. Ausín et al. (2010) and Kalli et al. (2011) introduced GARCH semiparametric models. Ausín et al. (2010) suggested modelling the error terms of the return equation with a Dirichlet process mixture of normals model. After fitting their semiparametric model to the Bombay Stock Exchange Index and the Hang Seng Index, they found evidence that their model better described the tail behaviour of the return distribution. Kalli et al. (2011) introduced an alternative semiparametric GARCH model where the error terms of the return equation are modelled using an infinite mixture of scaled uniform distributions. The empirical findings are similar to those discussed in Ausín et al. (2010). In stochastic volatility models, Jensen and Maheu (2010) and Delatola and Griffin (2011) introduce nonlinear and linear semiparametric SV models respectively using Dirichlet process mixture of normal models. Both models were shown to be better at capturing the tail behaviour of the returns than a simple SV model with a normal error distribution.

The scope of this paper is to extend the work of Nakajima and Omori (2009) using Bayesian nonparametric techniques. The error terms of the SV model will be flexibly modelled using the Dirichlet process mixture model (DPM) which allows for several of the stylized features, as discussed previously, to be captured. The flexibility of the DPM avoids the need to introduce extra parameters to capture some features of the return distribution. An alternative semiparametric SV model with leverage was introduced by Jensen and Maheu (2011) who used a bivariate DPM in the nonlinear SV model with leverage. In the empirical analysis of both Jacquier et al. (2004) and Nakajima and Omori (2009), there was evidence that their SV model with heavy tails and leverage fitted the examined data better than models based on the assumption of normality. These findings show that the commonly-made assumption of normality of error terms does not hold in many cases. However, both parametric models have computationally challenging schemes for updating both the the heavy tails and the leverage coefficient. On the contrary, the heavy tails in a semiparametric model can be captured by nonparametric techniques.

The paper is structured as follows: Section 2 describes our Bayesian nonparametric model with leverage (SVL-SPM), Section 3 reviews the sampling strategy for MCMC estimation of this model, Section 4 reports applications of the method to simulated and financial data examples (Microsoft asset prices and the S & P 500 index), and Section 5 concludes.

2. Semiparametric stochastic volatility model with leverage

This section presents a flexible version of the linear state-space representation of the SV model with leverage (SVL-SPM). The SVL-SPM extends the parametric model with leverage presented by Omori et al. (2007) which will be referred to as the SVL-PM. The next two subsections summarise the concepts of the SVL-PM and the Dirichlet process mixture (DPM) model respectively which will be used to build the SVL-SPM.

2.1. Parametric stochastic volatility model with leverage

A linear state-space representation of the non-linear SV model with leverage (SVL-PM) in (1) is derived by taking the logarithm of the squared returns. Hence, the SVL-PM is

$$y_t^\star = h_t + z_t,$$

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

where $y_t^\star = \log\,(y_t^2 + c)$, $c = 10^{-4}$ is the offset parameter and $z_t = \log \epsilon_t^2$. The correlation between ϵ_t and η_t can be accommodated using the following argument. Firstly, $p(z_t)$ can be accurately approximated by a mixture of normals

$$p(z_t) = \sum_{j=1}^{10} w_j \mathrm{N}(z_t|\mu_j, \sigma_j^2) \tag{2}$$

where $\mathrm{N}(x|m, \sigma^2)$ represents the density of a normal distribution with mean m and variance σ^2 (the values of $w_1, \ldots, w_{10}$, $\mu_1, \ldots, \mu_{10}$ and $\sigma_1^2, \ldots, \sigma_{10}^2$ are given in Omori et al. (2007)). It is useful to write this mixture in terms of allocation variables, s_t, which allocate an observation to a component of the mixture model

$$z_t|s_t = j \sim \mathrm{N}(\mu_j, \sigma_j^2), \qquad p(s_t = j) = w_j.$$

If we define $d_t = \mathrm{sign}(y_t) = \mathrm{I}(y_t > 0) - I(y_t \leq 0)$, then

$$\eta_t|z_t, d_t \sim \mathrm{N}\left(d_t \rho \sigma_\eta \exp\{z_t/2\}, \sigma_\eta^2(1-\rho^2)\right)$$

which Omori et al. (2007) suggest approximating (using a first-order Taylor series expansion of $\exp\{x\}$) by

$$\eta_t|z_t, d_t \sim \mathrm{N}\left(d_t \rho \sigma_\eta \exp\{\mu_{s_t}/2\}\left(a_{s_t} + b_{s_t}(z_t - \mu_{s_t})\right), \sigma_\eta^2(1-\rho^2)\right)$$

where $a_j = \exp\{\sigma_j^2/8\}$ and $b_j = a_j/2$ (their values can be found in Omori et al. (2007)). The full parametric model can be written

$$\left(\begin{array}{c} y_t^\star \\ h_t \end{array}\right)\Bigg| d_t, s_t, \rho, \sigma_\eta^2 = \left(\begin{array}{c} \mu_{s_t} + \sigma_{s_t}\epsilon_t^{(1)} \\ d_t \rho \sigma_\eta \exp\left\{\frac{\mu_{s_t}}{2}\right\}\left(a_{s_t} + b_{s_t}(y_t^\star - h_t - \mu_{s_t})\right) + \sigma_\eta\sqrt{1-\rho^2}\epsilon_t^{(2)} \end{array}\right)$$

where

$$\left(\begin{array}{c} \epsilon_t^{(1)} \\ \epsilon_t^{(2)} \end{array}\right) \sim \mathrm{N}_2\left(\mathbf{0}, \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right)\right).$$

2.2. Dirichlet process mixture model

As we mentioned in the introduction, the idea of using Bayesian nonparametric techniques to model the behaviour of financial data is not novel. We will follow much of the literature by assuming nonparametric forms for some distributions in our model to define a semiparametric model. Infinite component mixture models offer flexibility to capture many features of the conditional distribution of the return that cannot be obtained using a parametric model without introducing extra parameters.

The DPM of Ferguson (1983) and Lo (1984) is an infinite component mixture model which assumes that a sample $z_1, \ldots, z_n$ are independent and identically distributed from an unknown distribution. It builds on the Dirichlet process (Ferguson, 1973) which is a prior over probability distributions. The prior has two parameters: a mass parameter $M > 0$ and a probability distribution H. A Dirichlet process with these parameters will be written DP(M, H). A distribution G generated by a Dirichlet process are almost surely discrete and so can be written

$$G = \sum_{j=1}^{\infty} \zeta_j \delta_{\theta_j}$$

where $\theta_j \stackrel{i.i.d.}{\sim} H$ and δ_x is the Dirac delta function which places measure one on x. We will refer to $\zeta_1, \zeta_2, \ldots$ as the weights of a Dirichlet process with mass parameter M. The DPM assumes that the distributio of z_i is represented as

$$p(z_i) = \sum_{j=1}^{\infty} \zeta_j \mathrm{N}\left(z_i | \mu_j, \sigma_j^2\right)$$

where $\zeta_1, \zeta_2, \ldots$ are the weights generated by a Dirichlet process with mass parameter M and $(\mu_j, \sigma_j^2) \stackrel{i.i.d.}{\sim} H$. The prior can be considered the limit of a finite mixture model. Suppose that

$$p(z_i) = \sum_{j=1}^{K} \zeta_j \mathrm{N}\left(z_i | \mu_j, \sigma_j^2\right)$$

where $\zeta \sim \text{Di}(M/K, \ldots, M/K)$ and $(\mu_j, \sigma_j^2) \overset{i.i.d.}{\sim} H$. The limit of this model as $K \to \infty$ is the DPM. This use of an infinite component mixture model avoids the need to choose a value of K. The number of components with non-neglible weights (*i.e.* the j for which $\zeta_j > \epsilon$ for some small ϵ) is finite and so represents an effective value of the number of clusters. Its value is controlled by M with smaller values of M implying less non-neglible components on average *a priori*.

In this paper, the representation of the DPM as introduced by Griffin (2010) will be employed. This representation allows for a non-informative prior structure for location and scale parameters of the DPM. The model can be represented as:

$$p(z_t) = \sum_{j=1}^{\infty} \zeta_j \mathrm{N}\left(z_t | \mu_j^{'}, \alpha \sigma_z^2\right), \tag{3}$$

where $\zeta_1, \zeta_2, \ldots$ are the weights generated by a Dirichlet process with mass parameter M and $\mu_j^{'} \sim \mathrm{N}\left(\mu_0, (1-\alpha)\,\sigma_z^2\right)$. This model assumes that the component variances are constant. The parameter σ_z^2 is the prior variance of z_t and α plays the role of the smoothing parameter, $0 < \alpha < 1$. The prior expected distribution of z_t is normal with mean μ_0 and variance σ_z^2. The parameter α controls the variability of the distribution of z_t around its expectation. For values of α close to one, the distribution of z_t will be close to a normal distribution. While for values of α close to zero, the distribution of the errors z_t will be multimodal and far from normal. The hyperparameters can be treated as the location (μ_0), scale (σ_z) and smoothness (α) of the density. In fact, small values of α indicate a multimodal density while values bigger that 0.5 indicate a density closer to the normal distribution.

Griffin (2010) also introduced a model where component variances can differ but also allows the same non-informative prior structure as (3). This different component variance model is:

$$p(z_t) = \sum_{j=1}^{\infty} \zeta_j \mathrm{N}\left(z_t \middle| \mu_j^{'}, \alpha \frac{\zeta}{\mu_\zeta} \sigma_z^2\right)$$

where $\zeta_1, \zeta_2, \ldots$ are the weights generated by a Dirichlet process with mass parameter M and $\mu_j^{'} \sim \mathrm{N}\left(\mu_0, (1-\alpha)\,\sigma_z^2\right)$, where $\mu_\zeta = \mathrm{E}(\zeta)$, $\zeta \sim \mathrm{IG}(\zeta_0, 1)$ and

$\zeta_0 > 1$. Griffin (2010) finds little evidence that this model can estimate the the density of z_t better than the model with constant variance across components in (3) and, therefore, we use the simpler model with constant variances in our analysis.

2.3. The model (SVL-SPM)

The SVL-PM model approximates the distribution of the error terms in the return equation, which follow the $\log \chi_1^2$ distribution, by the ten component mixture model in (2). As mentioned earlier the assumption of normality is restrictive. Thus, an alternative approach pursued in this paper assumes that the error terms of the return equation are modelled nonparametrically using the representation of the DPM described in the previous section. Our proposed model, the SVL-SPM, is defined as

$$
\begin{aligned}
y_t^\star &= h_t + z_t,\\
h_{t+1} &= \mu + \phi\left(h_t - \mu\right) + d_t \rho \sigma_\eta \exp\left\{\frac{\mu_{s_t}^{'}}{2}\right\}\left[a^\star + b^\star\left(y_t - h_t - \mu_{s_t}^{'}\right)\right] + \sigma_\eta \sqrt{1-\rho^2}\epsilon_t^\star,\\
z_t|s_t &\sim \mathrm{N}(\mu_{s_t}', \alpha\sigma_z^2), \quad \text{and} \quad p(s_t = j) = \zeta_j
\end{aligned}
\tag{4}
$$

where $\zeta_1, \zeta_2, \ldots$ are the weights of the Dirichlet process with mass parameter M, $0 < \alpha < 1$, $\mu_j^{'} \overset{i.i.d.}{\sim} \mathrm{N}(0, (1-\alpha)\sigma_z^2)$ and $\epsilon_t^\star$ follows a standard normal distribution. The value of $a^\star$ and $b^\star$ are derived using the same argument as Omori et al. (2007), described in Section 2.1, and depend on the variance of each component. In this case, we have

$$
a^\star = \exp\left\{\frac{\alpha\sigma_z^2}{8}\right\} \qquad b^\star = 0.5a^\star. \tag{5}
$$

These are independent of the component indices which contrasts with the parametric model where a_j's and b_j's differs between components. The SVL-SPM is an extension of the semiparametric SV model (SV-SPM) of Delatola and Griffin (2011), which allowed for heavy tails in the return distribution but not a leverage effect. Similar to the SV-SPM, the information about the sign of the return y_t is lost, when taking the logarithm of the squares. The distribution of ϵ_t can only be recovered by making assumption about the conditional distribution of $d_t|z_t$ (that is the sign of ϵ_t given the log of the squared of ϵ_t). The simplest assumption is that ϵ_t is symmetric.

Similarly to other semiparametric models, see *e.g.* Bush and MacEachern (1996) and Jensen and Maheu (2010), an identifiability constraint must be imposed in order to be able to conduct inference. The constraint is $h_t^\star = h_t - \mu$,

$z_t^\star = z_t + \mu$ (or $\mu_j^{'\star} = \mu_j^{'} + \mu$) and $h_1^\star \sim \mathrm{N}\left(0, \frac{\sigma_\eta^2}{1-\phi^2}\right)$. Under this identifiability aint, the SVL-SPM becomes

$$y_t^\star = h_t^\star + z_t^\star \tag{6}$$

$$h_{t+1}^\star = \phi h_t^\star + d_t \rho \sigma_\eta \exp\left\{\frac{\mu_{s_t}^{'\star}}{2}\right\}\left[a^\star + b^\star\left(y_t - h_t^\star - \mu_{s_t}^{'\star}\right)\right] + \sigma_\eta\sqrt{1-\rho^2}\epsilon_t^\star.$$

This model no longer includes the intercept term μ which is subsumed into $z_t^\star$.

As in the SV-SPM model of Delatola and Griffin (2011), a connection between error terms of the return equation $z_t^\star$ and the error terms of the return equation of the nonlinear model ϵ_t can be obtained. The variance of ϵ_t can be approximated by

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

and the kurtosis of ϵ_t can be approximated by

$$\mathrm{K}(\epsilon_t) \approx \mathrm{Var}(z_t^\star) - 1.$$

The mean and variance of $z_t^\star$ can be approximated using the output from the MCMC algorithm needed to fit the SVL-SPM model. The mean of $z_t^\star$ conditional on the output from one iteration of the MCMC algorithm is

$$\mathrm{E}[z_t^\star|\psi] = \sum_{j=1}^{k} \frac{n_j}{n+M}\mu_j^{'} + \frac{M}{M+n}\mu_0, \tag{7}$$

where n is the number of observations, the observations are allocated to the K clusters, n_j is the number of observations allocated to the j^{th} cluster, and $\psi = (K, n_1, \ldots, n_K, \mu_1', \ldots, \mu_k', \mu_0, a, \sigma_z^2, M)$. The parameter μ is not directly estimated in our model. It is useful to define the value $\mu = \mathrm{E}[z^\star] + 1.2704$ which would be the value of μ in the parametric model if ϵ_t is standard normally distributed. The variance is

$$\mathrm{V}[z_t^\star|\psi] = \frac{\sum_{j=1}^{k}\left(n_j\left(\alpha\sigma_z^2 + \mu_j^{'2}\right)\right) + M\left(\mu_0^2 + \sigma_z^2\right)}{n+M} - \mathrm{E}[z_t^\star|\psi]^2 \tag{8}$$

Unlike Delatola and Griffin (2011), the value of α is inferred from the data and given a uniform prior distribution (in our experience, the SVL-SPM does not suffer from small draws of α, which lead to poor mixing as in the model of Delatola and Griffin (2011)). The priors for the other parameters of the SVL-SPM model are, respectively

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

where IG(a,b) is an inverse Gamma distribution with mean (if $a > 1$) $\frac{b}{a-1}$ and variance (if $a > 2$) $\frac{b^2}{(a-1)(a-2)}$ and $\mathrm{N}_{[a,b]}(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2 restricted to the interval (a, b). The truncated prior for ϕ imposes stationarity on the log-volatility process. The mass parameter of the Dirichlet process, M, has a prior suggested by Escobar and West (1995), which is an exponential distribution with mean 2, $M \sim \mathrm{Exp}(2)$.

3. MCMC Algorithm for SVL-SPM

This section is devoted to presenting the steps of the MCMC algorithm for the SVL-SPM. The representation of the SVL-SPM as a linear state-space model in (6) will be used to conduct inference. The model in (4) is a non-conjugate Dirichlet process mixture model and so some important modifications are needed to the sampling scheme for the SV-SPM described in Delatola and Griffin (2011).

As is common with MCMC schemes for mixture models, allocation variables s_t are introduced as in (4). Let $\mathbf{y}^\star = \{y_t^\star\}_{t=1}^n$, $\mathbf{h}^\star = \{h_t^\star\}_{t=1}^n$, $\boldsymbol{\mu}'^\star = \left\{\mu_j'^\star\right\}_{j=1}^n$ and $\mathbf{s} = \{s_t\}_{t=1}^n$. It is assumed that there are K distinct values of $s_1, \ldots, s_n$, and that there are n_j observations for which $s_t = j$. The steps for the MCMC algorithm are the following:

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

3.1. Updating the log-volatilities

The sampling scheme to update the log-volatilities is in line with the forward filtering backward sampling (FFBS) algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The equations of the FFBS are based on the definition of the state-space model as given by De Jong (1991).

Definition 1. *Let a random vector* $\mathbf{y}^\star = (y_1^\star, y_2^\star, \ldots, y_n^\star)$ *be generated from a state space model for* $t = 1, \ldots, n$*, then*

$$y_t^\star = c_t + Z_t h_t^\star + G_t u_t \qquad \textit{(Observation Equation)}, \tag{9}$$

where $t = 0, \ldots, n$*, and*

$$h_{t+1}^\star = f_t + T_t h_t^\star + H_t u_t \qquad \textit{(System Equation)}. \tag{10}$$

The assumptions made are:

- $u_t \sim N(0, 1)$,
- $h_0^\star = 0$,
- c_t *and* f_t *are assumed known.*

The FFBS can be used for the SVL-SPM with

$$f_t = d_t \rho \sigma_\eta \exp\{\mu_{s_t}^{'\star}/2\} \left[a^\star + b^\star \left(y_t^\star - \mu_{s_t}^{'\star}\right)\right],$$

$G = \alpha \sigma_z^2$, $H_t = \sigma_\eta \sqrt{1 - \rho^2}$ and $T_t = \left(\phi - d_t \rho \sigma_\eta \exp\{\mu_{s_t}^{'\star}/2\} b^\star\right)$.

3.2. Updating the allocation variables

The allocation variables s are updated using algorithm 8 of Neal (2000). The allocation variables appear both in the log-return equation and in the log-volatility equation. The parameter s_t is updated conditional on

$$s_{-t} = (s_1, \ldots, s_{t-1}, s_{t+1}, \ldots, s_n).$$

Let K_{-t} be the number of distinct values in s_{-t} and let n_j^- be the number of observations in s_{-t} for which $s_k = j$. A number, m, of empty clusters are introduced in the algorithm. Typically m is chosen to be small ($m = 3$ was used in the examples in this paper). The full conditional distribution of s_t is a discrete distribution with $K+m$ possible values. The probability that $s_t = j$ is

$$p_j \propto w_j^\star \exp\left\{-\frac{\left(y_t^\star - h_t^\star - \mu_j^{'}\right)^2}{2\alpha\sigma_z^2}\right\}$$

$$\times \exp\left\{\frac{\left(h_{t+1}^\star - \phi h_t^\star - d_t\rho\sigma_\eta \exp\left\{\frac{\mu_j^{'\star}}{2}\right\}\left[a^\star + b^\star\left(y_t^\star - h_t^\star - \mu_j^{'\star}\right)\right]\right)^2}{2\sigma_\eta^2\left(1-\rho^2\right)}\right\},$$

for $j = 1, \ldots, K+m$ where the allocation weight $w_j^\star$ is

$$w_j^\star = \begin{cases} \frac{n_j}{(M+n-1)\sqrt{\alpha\sigma_z^2\sigma_\eta^2(1-\rho^2)}}, & \text{for } j = 1, \ldots, K \\ \frac{M/m}{(M+n-1)\sqrt{\alpha\sigma_z^2\sigma_\eta^2(1-\rho^2)}}, & \text{for } j = k+1, \ldots, K+m. \end{cases}$$

The above formulae are valid only for $t = 1, \ldots, (n-1)$. Omori et al. (2007) showed in the parametric counterpart of the leverage model that the allocation probability for the observation at time $t = n$ is different from times $t = 1, \ldots, n-1$. Thus, p_j for $t = n$ should be

$$p_j \propto w_j^\star \exp\left\{-\frac{\left(y_n^\star - h_n^\star - \mu_j^{'\star}\right)^2}{2\alpha\sigma_z^2}\right\},$$

where $w_j^\star$ is defined as before.

3.3. Updating the Location Parameters

In this step, a Metropolis-Hastings independence sampler is employed to draw $\boldsymbol{\mu}'$. For each component, the proposal is sampled from a normal distribution

$$\mu_j^{'\star}\text{can} \sim \mathrm{N}\left(\frac{\frac{\sum_{\{t|s_t=j\}}(y_t^\star - h_t^\star)}{\alpha} + \frac{\mu_0}{1-\alpha}}{\frac{n_j}{\alpha} + \frac{1}{1-\alpha}}, \frac{\sigma_z^2}{\frac{n_j}{\alpha} + \frac{1}{1-\alpha}}\right).$$

Each candidate is accepted with probability

$$p = \min\left(1, \frac{f\left(\mu_j^{'\star}\text{can}|\mu_j^{'\star}\right)}{f\left(\mu_j^{'\star}|\mu_j^{'\star}\text{can}\right)}\right),$$

where $f\left(\mu_j^{'\star}|\mu_j^{'\star}\text{can}\right) = \exp\left\{\frac{\sum_{\{t|s_t=j\}}\left(h_{t+1}^\star - \phi h_t^\star - d_t\rho\sigma_\eta\exp\left\{\frac{\mu_j^{'\star}}{2}\right\}\left[a^\star + b^\star\left(y_t - h_t^\star - \mu_j^{'\star}\right)\right]\right)^2}{2\sigma_\eta^2(1-\rho^2)}\right\}.$

3.4. Updating μ_0

The full conditional for μ_0 follows a normal distribution

$$\mu_0 \sim \mathrm{N}\left(\frac{\sum_{i=1}^K \mu_i^{'\star}}{K}, \frac{(1-\alpha)\sigma_z^2}{K}\right).$$

3.5. Updating the Smoothing Parameter

Initial simulations studies were carried out using the representation of the DPM as shown in (3). However, using this representation caused mixing problems in the algorithm. Specifically, the smoothing parameter α and the variance parameter σ_z^2 were negatively correlated. Griffin (2006) introduced a representation of the DPM allowing joint updating of α and σ_z^2. The representation of the DPM as suggested by Griffin (2006) is

$$z_t|\mu_j^{'} \sim \mathrm{N}\left(\mu_j^{'}, \sigma_z^{2\star}\right) \tag{11}$$

and

$$\mu_j^{'} \sim \mathrm{N}\left(\mu_0, \sigma_z^{2'}\right).$$

This is a reparameterization of α and σ_z^2 in the SVL-SPM with $\sigma_z^{2\star} = \alpha\sigma_z^2$ and $\sigma_z^{2'} = (1-\alpha)\sigma_z^2$.

The parameters (α, σ_z^2) will be updated simultaneously using an independence Metropolis-Hastings sampling step based on the reparameterization of Griffin (2006). The proposal is given by Griffin (2006) and can be rejection sampled. For purposes of completeness it is also presented here. Suppose that $\widehat{\alpha}$ is the current value of α, a value is simulated from the proposal by sampling from the rejection envelope

$$\sigma_z^{2\star\text{can}} \sim \mathrm{IG}\left(\frac{n}{2} + \widehat{\alpha} - (1-\widehat{\alpha}), \frac{1}{2}\sum_{i=1}^{n}\left(y_t^\star - h_t^\star - \mu_{s_t^\star}^{'\star}\right)^2\right)$$

and

$$\sigma_z^{2'\text{can}} \sim \mathrm{IG}\left(\frac{K}{2} - \widehat{\alpha} + (1-\widehat{\alpha}), \frac{1}{2}\sum_{i=1}^{K}\left(\mu_i^{'\star} - \mu_0\right)^2\right).$$

The sampled value is accepted as a draw from the proposal with probability

$$\frac{1}{\sigma_z^{2\star\text{can}} + \sigma_z^{2'\text{can}}}^2 \left(\frac{\sigma_z^{2\star\text{can}}}{\widehat{\alpha}}\right)^{2\widehat{\alpha}} \left(\frac{\sigma_z^{2'\text{can}}}{1-\widehat{\alpha}}\right)^{2(1-\widehat{\alpha})}.$$

The acceptance probability for the Metropolis-Hastings sampling step is

$$p = \min\left(1, \frac{f\left(\sigma_z^{2\prime}\text{can}, \sigma_z^{2\star}\text{can} | \sigma_z^{2\prime}, \sigma_z^{2\star}\right)}{f\left(\sigma_z^{2\prime}, \sigma_z^{2\star} | \sigma_z^{2\prime}\text{can}, \sigma_z^{2\star}\text{can}\right)}\right),$$

where

$$f\left(\sigma_z^{2\prime}, \sigma_z^{2\star}\right) = \left(\sigma_z^{2\star}\right)^{-n/2}\left(\sigma_z^{2\prime}\right)^{-K/2} \exp\left\{-\frac{\sum_{j=1}^{K}\left(\mu_j^{\prime\star} - \mu_0\right)^2}{2\sigma_z^{2\prime}} - \frac{\sum_{t=1}^{n}\left(y_t^\star - h_t^\star - \mu_{s_t^\star}^{\prime\star}\right)^2}{2\sigma_z^{2\star}}\right\}$$

$$\times \exp\left\{\frac{\sum_{t=1}^{n-1}\left(h_{t+1}^\star - \phi h_t^\star - d_t\rho\sigma_\eta \exp\left\{\frac{\mu_{s_t^\star}^{\prime\star}}{2}\right\}\left[\exp\left\{\sigma_z^{2\star}/8\right\} + 0.5\exp\left\{\sigma_z^{2\star}/8\right\}\left(y_t^\star - h_t^\star - \mu_{s_t^\star}^{\prime\star}\right)\right]\right)^2}{2\sigma_\eta^2(1-\rho^2)}\right\}.$$

3.6. Updating the Mass Parameter

The mass parameter M can be updated uisng the sampling scheme suggested by Escobar and West (1995).

3.7. Updating the Volatility Parameters

The parameters of the volatility equation are updated using adaptive Metropolis-Hastings random walks. The likelihood $f\left(\mathbf{y}^\star | \{d_t\}_{t=1}^n, \mathbf{s}, \phi, \rho, \sigma_\eta^2\right)$ for sampling the parameters ϕ, ρ and σ_η^2 of the volatility equation is

$$f(\mathbf{y}^\star) \propto \left(1-\rho^2\right)^{-\frac{n-1}{2}}\left(1-\phi^2\right)^{-\frac{1}{2}}\left(\sigma_\eta^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{h_1^{2\star}\left(1-\phi^2\right)}{2\sigma_\eta^2}\right\}\times$$

$$\exp\left\{-\frac{\sum_{t=1}^{n-1}\left(h_{t+1}^\star - \phi h_t^\star - d_t\rho\sigma_\eta \exp\left\{\mu_{s_t^\star}^{\prime\star}/2\right\}\left[a^\star + b^\star\left(y_t^\star - h_t^\star - \mu_{s_t^\star}^{\prime\star}\right)\right]\right)^2}{2\sigma_\eta^2\left(1-\rho^2\right)}\right\}.$$

Each parameter of the volatility equation is updated individually conditional on the other parameters using the adaptive Metropolis-Hastings random walk algorithm of Atchadé and Rosenthal (2005). We use the transformed parameters $z_\phi = \log\phi - \log(1-\phi)$, $z_\rho = \log\rho - \log(1-\rho)$ and $z_\eta = \log\sigma_\eta^2$. These transformed parameters are updated using a usual Metropolis-Hastings sampling step. Let σ_ρ^2, σ_ϕ^2 and $\sigma_{\sigma_\eta^2}^2$ be the variance of a normal increment in the random walk for

z_ϕ, z_ρ and z_η respectively. The variance of the increment is updated at each iteration of the algorithm. For example, the variance term for z_ρ would be updated by

$$\sigma_\rho^2 = \sigma_\rho^2 + \frac{1}{j^{0.5}}\left(p - 0.3\right),$$

where j is the iteration at the current point, p is the acceptance rate for updating ρ and 0.3 is the acceptance rate. Thus, the scaling parameter is adapted at each iteration.

4. Examples

This section describes the results of fitting the semiparametric SV model with leverage (SVL-SPM) to both simulated and empirical data sets. The empirical data sets were the asset returns of Microsoft on the NASDAQ and the S&P 500 stock index of the New York Stock Exchange. The predictive performance of the model was compared to the performace of two other models: the SVL-PM and the semiparametric SV model (SV-SPM) of Delatola and Griffin (2011).

The codes for all three models were written in MatLab and run using two quad core Xeon 2.53Ghz CPUs. Each sampler was run with a burn-in period of $100\,000$ iterations. After this period, the code was run for an additional $50\,000$ iterations storing every 5^{th} draw. Table 1 shows the CPU times in seconds for the two models with leverage (SVL-SPM and SVL-PM). The SVL-SPM takes between 20% and 60% longer than the SVL-PM over the three data sets.

	SVL-SPM	SVL-PM
Simulated	6150	4985
Microsoft	9861	6148
S&P 500	10719	7942

Table 1: CPU times (in seconds) for the SVL-SPM and SVL-PM. The CPU times were calculated when running the samplers for $10\,000$ iterations in an 2GHz Intel Core 2 Duo processor.

The predictive performance of the models were assessed by both the average log-predictive score for one-step ahead predictions (LPS) of Kim et al. (1998) and the conditional likelihood score (CL) of Diks et al. (2011). Both criteria are proper scoring rules and so can be used to compare the predictions of the models. The model parameters are $\theta = (\phi, \sigma_\eta^2, \rho, F)$ where F is the distribution of z_t^*. 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 the posterior mean of the model parameters. The one-step ahead predictive density is given by

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

The predictive distribution was accurately and efficiently approximated using a sequential Monte Carlo algorithm. Smaller values of the LPS indicate a model giving better one-step ahead predictions. A drawback with the LPS is that all predictions contribute equally to the score. In practice, we may be more interested in the ability of models to predict extreme returns than returns in the centre of their distribution. Models which predict these events better are typically more accurately modelling the tails of the return distribution. By definition, these events are relatively rare and so the difference may not be clearly shown by the LPS method. Diks et al. (2011) discussed proper scoring rules and tests of predictive performance when the scoring rule concentrates on a subset of the observations. Their conditional likelihood method was used to give a scoring rule which concentrates on extreme events (and so the tails of the return distribution). We use the conditional likelihood score

$$\text{CL} = -\frac{1}{\sum_{i=1}^{T} I\left(y_i^\star \geq z_b\right)}\sum_{i=1}^{T} I\left(y_i^\star \geq z_b\right)\log\left(\frac{p\left(y_i^\star \left|\mathbf{y}_{1:(i-1)}^\star, \hat{\theta}\right.\right)}{\int_{z_b}^{\infty} p\left(y_j^\star \left|\mathbf{y}_{1:(j-1)}^\star, \hat{\theta}\right.\right)\, dy_j^\star}\right),$$

where z_b represents the upper $100\alpha\%$ point of the empirical distribution of the squared returns. Only predictions for returns with absolute value in the upper $100\alpha\%$ of their empirical distribution are included in the score and the score replaces the full distribution used by the LPS with the conditional distribution given that the return is above z_b. Again, the score with the smallest value indicates the better fit. Diks et al. (2011) showed that this type of score can effectively discriminate between the predictive ability of models in the tails.

4.1. Simulated Example

The three models were initially compared using data generated to have both leverage and heavy tails. The data had 3000 observations and were simulated from

	True	SVL-SPM	SVL-PM	SV-SPM
ϕ	0.97	0.956 *(0.923,0.971)*	0.903 *(0.883,0.948)*	0.956 *(0.929,0.975)*
μ	0.00	0.124 *(0.000,0.240)*	0.163 *(0.047,0.328)*	-0.716 *(-1.038,-0.392)*
σ_η	0.150	0.173 *(0123,0.226)*	0.322 *(0.228,0.349)*	0.194 *(0.142,0.255)*
ρ	-0.600	-0.576 *(-0.713,-0.427)*	-0.422 *(-0.548,-0.277)*	
α		0.178 *(0.072,0.345)*		0.001 *(0.000, 0.005)*
σ^2		4.771 *(4.577,4.928)*		5.646 *(4.776,10.371)*
M		1.117 *(0.321,2.872)*		0.456 *(0.131, 1.149)*
k		11 *(5,23)*		2 *(2,4)*

Table 2: Simulated data: Posterior medians and 95% credible intervals for the SVL-SPM, the SVL-PM and the SV-SPM with $c = 10^{-4}$.

the following model:

$$y_t = \exp\left(h_t/2\right)\epsilon_t\sqrt{\lambda_t^{-1}}, \qquad \lambda_t \sim \text{Ga}(7/2, 7/2),$$

and

$$h_{t+1} = 0.97\, h_t + \eta_t$$

where

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.6\times\sqrt{0.0225} \\ -0.6\times\sqrt{0.0225} & 0.0225 \end{pmatrix}\right).$$

The model implies that the marginal distribution of ϵ_t is a Student t distribution with 7 degrees of freedom, which has substantially heavier tails than a normal distribution. The leverage effect, ρ, is -0.6.

Table 2 presents the posterior medians and 95% credible intervals (CI) of the model parameters under the parametric and semiparametric models. The parameter estimates for the SVL-PM were based on the assumption that the error terms of the return equation follow a normal distribution with the estimates reweighted to take into account the error of approximationg the $\log\chi_1^2$ distribution by a mixture of normals. The SVL-PM clearly failed to estimate the parameters well and to capture the behaviour of the data. The persistence parameter ϕ and the leverage effect ρ were underestimated while μ and σ_η were overestimated. The model also tended to infer much larger increase in volatility in response to large values of ϵ_t and so underestimates ϕ and overestimates σ_η. The SV-SPM tends to underestimate μ since the correlation between ϵ_t and η_t is not included in the model. The SVL-SPM performs much better than both models with the true values of all

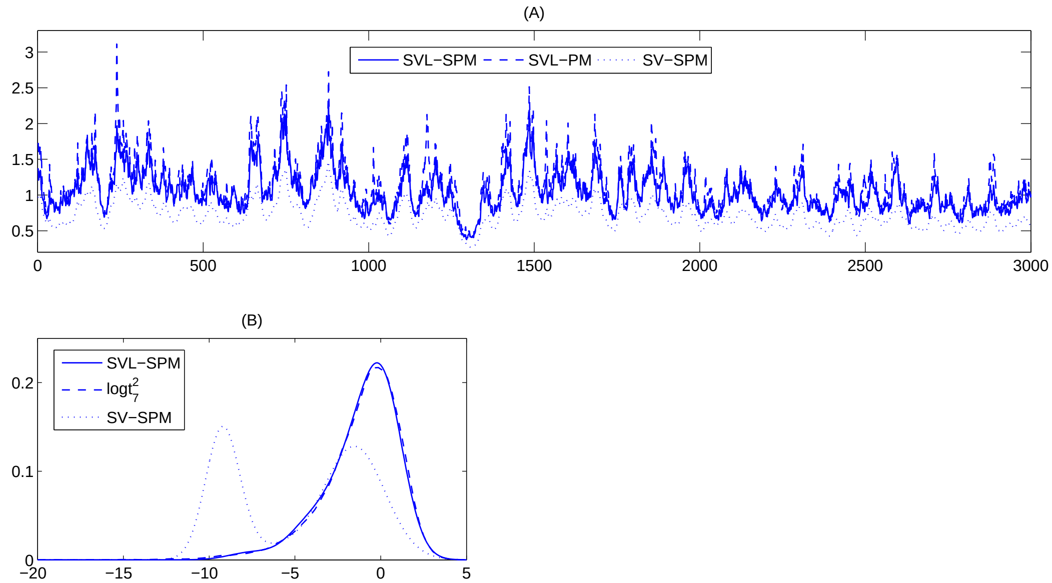

Figure 1: Simulated data: Posterior mean volatilities for the SVL-SPM, the SVL-PM and the SV-SPM (Panel A); Posterior mean densities for SVL-SPM and SV-SPM (Panel B).

parameters within the 95% credible interval. The kurtosis of the Student t distribution with 7 degrees of freedom is 5. The variance σ^2 can be used to approximate $\mathrm{K}(\epsilon_t)$ under the SVL-SPM and gives a value of the kurtosis around 4. The approximation underestimates the kurtosis here but it does illustrate the useful of this method as an indication of the heaviness of the tails of the return distribution.

Figure 1 depicts posterior estimates of the volatilities and the distribution of $z_t^\star$ for the different models. The posterior median for the volatilities under the SVL-PM had more peaks than the SVL-SPM which accommodated the extreme price movements generated by the Student t-distribution. This illustrates the effect of the underestimated persistence and the overestimated volatility of the log-volatilities (the parameters ϕ and σ_η^2) on the estimated volatilities. The SV-SPM tended to systematically underestimate volatility which directly links to the underestimation of the mean of the log-volatility (the parameter μ) under this model. On the contrary, the SVL-SPM gives smoother estimates of the volatility. The estimated density of $z_t^\star$ under the SVL-SPM gave a good approximation to the $\log t_7^2$ distribution. The SV-SPM model introduced an extra mode in this distribution around -10 and so allowed for the lack of a leverage effect by introducing a large number of very small returns.

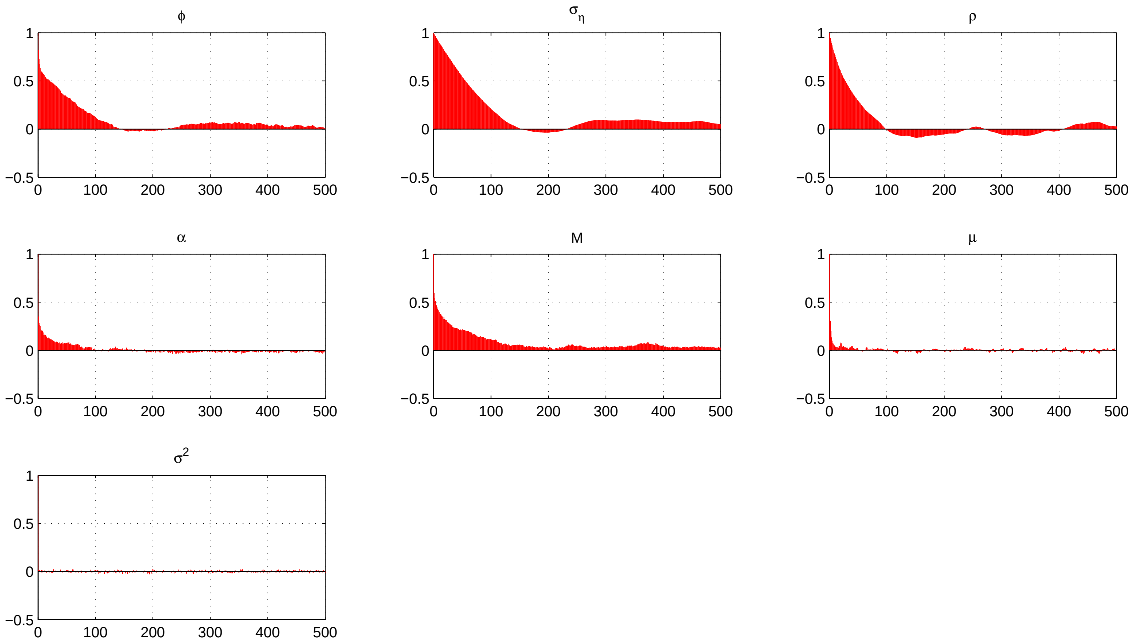

Figure 2: Simulated data: Autocorrelation plot for SVL-SPM.

Figure 2 shows the mixing of the different parameters of the SVL-SPM. Most parameter mixed well but there were some indications that the volatility parameters (ϕ, ρ and σ_η^2) suffered from relatively slow mixing in this example.

	SVL-SPM	SVL-PM	SV-SPM
LPS	2.22	2.23	2.58
CL	0.37	1.19	0.45

Table 3: Simulated data: Log-predictive scores and conditional likelihood score for the SVL-SPM, the SVL-PM and the SVL-PM for $c = 10^{-4}$.

Finally, Table 3 compares the fits of the the models. Both models with leverage had similar LPS's which were much smaller than the score with the SV-SPM. Therefore, it was hard to distinguish which model with leverage had a better overall fit. The CL score for both semiparametric models were much smaller than the score for the SVL-PM and indicated that these models were better at capturing the tail behaviour of the conditional distribution of returns. Overall, the SVL-SPM was better than the other models according to both scores.

4.2. Microsoft Asset Prices

The daily price series of Microsoft from January 4, 1993 to December 31, 2008, which has $n = 4030$ data points, were analyzed using the three models. The data were expressed as compounded returns in percentages and can be seen in Panel A of Figure 3.

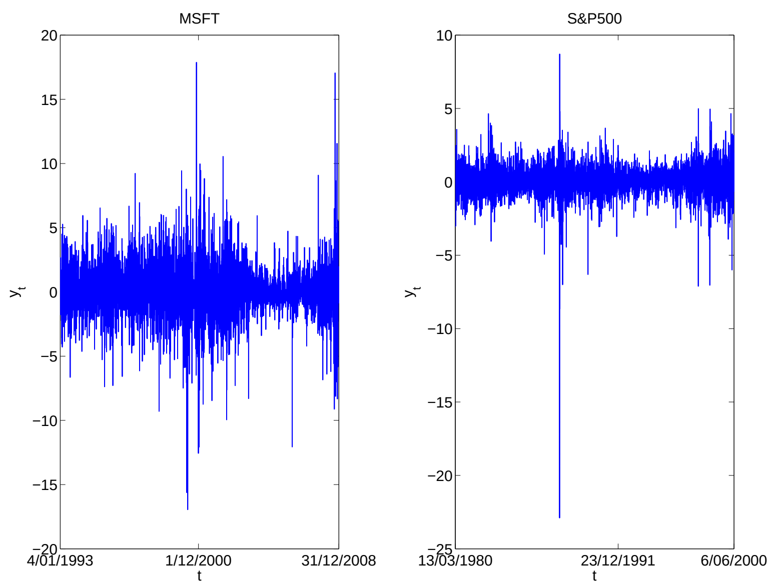

Figure 3: The log returns of (a) Microsoft asset prices and (b) the S&P500 index.

Table 4 presents the posterior estimates of the parameters under the different models. The posterior median of ϕ was larger for the SVL-SPM and SV-SPM than for the SVL-PM. The posterior estimate of the leverage, ρ, for the SVL-SPM was negative and zero was not contained in the 95% credible interval. In contrast, the SVL-PM had a larger posterior median of ρ which was positive with zero included in the 95% credible interval. This suggested evidence of no leverage effect under the SVL-PM, unlike the SVL-SPM which provided evidence of a leverage effect. In the case of σ_η, the posterior median was much smaller under the SVL-SPM than both SV-SPM and SVL-PM. These estimates reflected the misspecification of these models for these data. The SVL-PM could not capture the distribution

	SVL-SPM	SVL-PM	SV-SPM
ϕ	0.998 *(0.997,0.998)*	0.985 *(0.974,0.992)*	0.997 *(0.995,0.999)*
μ	1.813 *(1.069,2.508)*	1.142 *(0.783,1.524)*	1.255 *(-0.068,3.233)*
σ_η	0.042 *(0.035,0.050)*	0.165 *(0.126,0.208)*	0.070 *(0.067,0.074)*
ρ	-0.161 *(-0.273,-0.080)*	0.025 *(-0.176,0.196)*	
α	0.002 *(0.001,0.003)*		0.000 *(0.000,0.000)*
σ^2	6.262 *(6.221,6.300)*		5.799 *(5.509,6.054)*
M	7.115 *(4.396,11.014)*		42.695 *(30.033,62.25)*
k	59 *(43,81)*		197 *(155,259)*

Table 4: Microsoft asset prices: Posterior medians and 95% credible intervals for the SVL-SPM, the SVL-PM and the SV-SPM for $c = 10^{-4}$.

of the data and so σ_η is overestimated in a similar way to the simulated example. One reason the posterior median of σ_η was larger under the SV-SPM than the SVL-SPM was probably that the model could not capture the correlation between the errors ϵ_t and η_t which led to larger estimates of the variabilities. The posterior median of σ^2 suggested that the distribution of ϵ_t was heavy tailed. Interestingly, the mass parameter of the DPM differed significantly between the two semiparametric models.

The posterior estimates of the volatility and the predictive distribution of the SVL-SPM, the SVL-PM and the SV-PM can be seen in Figure 4. The posterior median of the volatility was much smoother under the SVL-SPM. This was consistent with a much smaller value of the posterior median of the volatility of the log-volatility, σ_η, under this model. Similarly, the posterior medians of the volatilities under the SV-SPM were smoother in comparison to the SVL-PM. The estimated density was multimodal for both semiparametric models. The posterior median of the number of components, k, were 59 and 197 for the SVL-SPM and the SV-SPM respectively.

	SVL-SPM	SVL-PM	SV-SPM
LPS	2.17	2.17	2.09
CL	0.42	0.68	0.50

Table 5: Microsoft asset prices: Log-predictive scores and conditional likelihood score for the SVL-SPM, the SVL-PM and the SV-SPM for $c = 10^{-4}$.

The LPS and CL score for the three models are reported in in Table 5. The SV-SPM outperformed the two models with leverage under the LPS and so pro-

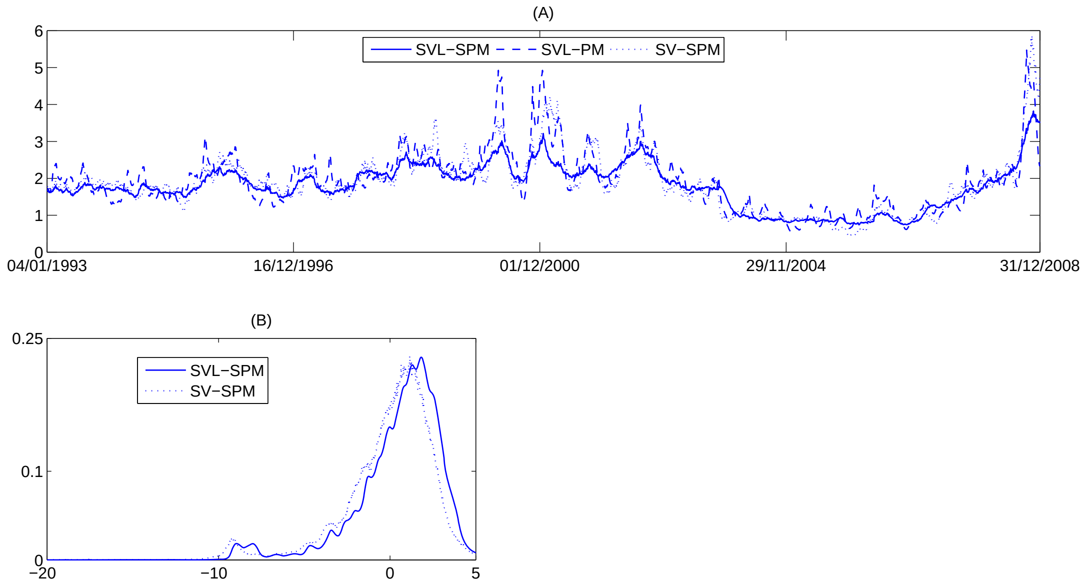

Figure 4: Microsoft asset price: Posterior mean volatilities for the SVL-SPM, the SVL-PM and the SV-SPM (Panel A); Posterior mean densities for SVL-SPM and SV-SPM (Panel B).

vided better overall predictions. However, the SVL-SPM had a smaller CL score than the SV-SPM (with both semiparametric models outperforming the SVL-PM) which indicated that this model gave better prediction of extreme returns.

4.3. S&P 500 Index

The compounded returns of the S&P 500 index were taken from March 13, 1980 to June 6, 2000 which led to 5136 data points. The returns are plotted in Panel B of Figure 3

Table 6 reports the parameter estimates under the SVL-SPM, the SVL-PM and the SV-SPM. Many of the findings were similar to the ones using the Microsoft asset price data. For example, the posterior median for the persistence parameter ϕ was much larger under the semiparametric models than the SVL-PM with the 95% credible intervals under the semiparametric models not crossing the interval for the SVL-PM. The same behaviour was observed for the volatility of the log-volatility, the posterior median was two times larger under the SVL-PM than under the semiparametric models. These results coincide with the ones found by Delatola and Griffin (2011) using the same data. Furthermore, the posterior me-

	SVL-SPM	SVL-PM	SV-SPM
ϕ	0.992 *(0.986,0.996)*	0.976 *(0.966,0.986)*	0.994 *(0.989,0.998)*
μ	-0.114 *(-0.396,0.305)*	-0.186 *(-0.363,0.037)*	-0.102 *(-0.568,0.370)*
σ_η	0.086 *(0.067,0.111)*	0.166 *(0.126,0.208)*	0.073 *(0.058,0.091)*
ρ	-0.477 *(-0.598,-0.329)*	-0.542 *(-0.568,-0.346)*	
α	0.080 *(0.035,0.151)*		0.001 *(0.000,0.003)*
σ^2	4.820 *(4.714,4.928*		4.843 *(3.945,7.161)*
M	1.518 *(0.508,3.512)*		0.685 *(0.236,1.533)*
k	15 *(7,29)*		4 *(3,8)*

Table 6: S&P500 index: Posterior medians and 95% credible intervals for the SVL-SPM, the SV-PM and the SV-SPM for $c = 10^{-4}$.

dian of the smoothing parameter α was much bigger under the SVL-SPM with this dataset than with the Microsoft asset prices, suggesting a smoother density for this data. The posterior median for α was smaller under the SV-SPM than that under the SVL-SPM. The leverage effect was much stronger in this data than the Microsoft asset price data with a posterior median of -0.54 under the parametric model and -0.48 under the semiparametric model with leverage.

Figure 5 compares the posterior estimates of the log-volatility using the different models. The figure in Panel A shows that the estimated volatility was much smoother under the semiparametric models than the SVL-PM, which was prone to introduce large peaks in response to large absolute returns. The SV-SPM gave much smoother volatility estimates than the SVL-SPM. The estimated density of $z_t^\star$ was bimodal under the SV-SPM but unimodal under the SVL-SPM. This corroborates the finding of the simulated example where the the SV-SPM model introduced an extra component around -10 in the estimated density to allow for the lack of a leverage effect term.

	SVL-SPM	SVL-PM	SV-SPM
LPS	2.12	2.12	1.88
CL	0.44	0.53	0.56

Table 7: S&P500 index: Log-predictive scores and conditional likelihood score for the SVL-SPM, the SVL-PM and the SV-SPM for $c = 10^{-4}$.

The LPS and CL score were calculated and are shown in Table 7. The SV-SPM gave a lower LPS than both the SVL-SPM and the SVL-PM and so indicated a better overall. However, the SVL-SPM was the best performing model under the

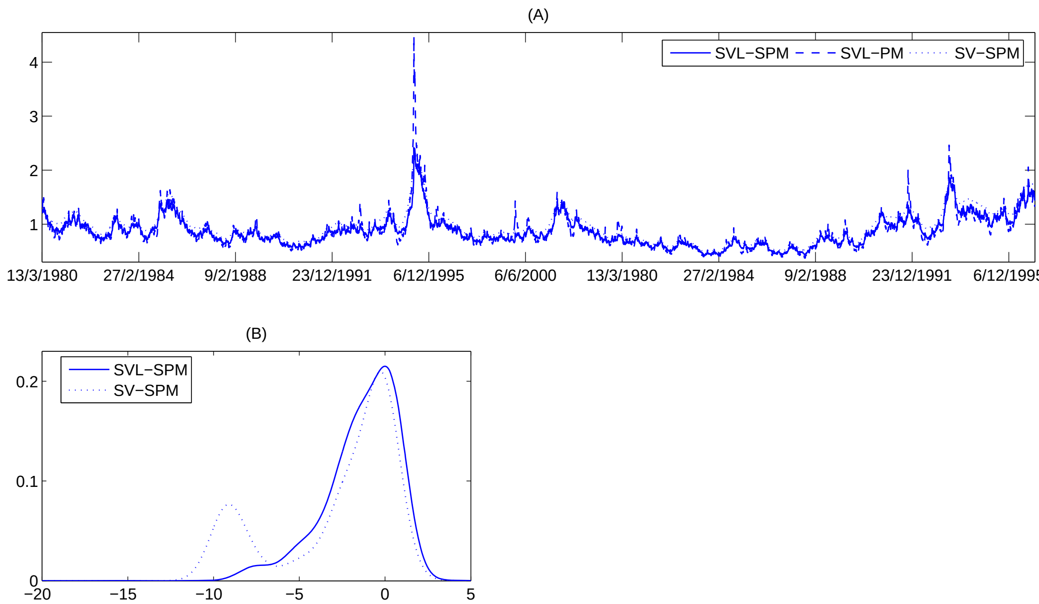

Figure 5: S&P500 index: Posterior mean volatilities for the SVL-SPM, the SVL-PM and the SV-SPM (Panel A); Posterior mean densities of SVL-SPM and SVL-SPM (Panel B).

CL score indicating that it could better fit the data in the tail of the conditional return distribution.

5. Discussion

This paper has discussed including a leverage effect in a Bayesian semiparametric model for volatility, which also allows for heavy tails. The application of this model to financial data shows that these changes to the model can have a considerable effect on the estimates of the leverage effect and lead to more sensible results. The semiparametric model without a leverage (SV-SPM) tends to include an extra mode in the conditional distribution of the squared returns if the data exhibit a strong leverage effect. The model can be fitted efficiently by updating the volatilities as a single block using standard forwards-backwards algorithms for state space models. However, although model is semiparametric, it assumes that the return distribution is symmetric, that the relationship between the innovations of the return and volatility equations are linearly related and that the innovations of the volatility equation are normally distributed. Future work will address these limitations of the model.

Acknowledgements

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.

References

Atchadé, Y. F. and J. S. Rosenthal (2005). On Adaptive Markov Chain Monte Carlo Algorithms. *Bernoulli 11*(5), 815–828.

Ausín, C., P. Galeano, and P. Ghosh (2010). A Semiparametric Bayesian Approach to the Analysis of Financial Time Series with Applications to Value at Risk Estimation. Technical report, Universidad Carlos III, Departamento de Estadística y Econometría.

Black, F. (1976). Studies of Stock Market Volatility Changes. In *Proceedings of the American Statistical Association, Business and Economic Studies Section*, pp. 177–181.

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

Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika 81*, 541–553.

De Jong, P. (1991). The Diffuse Kalman Filter. *The Annals of Statistics 19*(2), 1073–1083.

Delatola, E.-I. and J. E. Griffin (2011). Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility. *Bayesian Analysis* (6), 901–926.

Diks, C., V. Panchenko, and D. van Dijk (2011, August). Likelihood-Based Scoring Rules for Comparing Density Forecasts in Tails. *Journal of Econometrics 163*(2), 215–230.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*, 577–588.

Ferguson, T. S. (1973). Bayesian Analysis of some Nonparametric Problems. *Annals of Statistics 1*(2), 209–230.

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

Frühwirth-Schnatter, S. (1994). Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis 15*, 183–202.

Griffin, J. E. (2006). On the Bayesian Analysis of Species Sampling Mixture Models for Density Estimation. Technical report, Department of Statistics, University of Warwick.

Griffin, J. E. (2010). Default Priors for Density Estimation with Mixture Models. *Bayesian Analysis 5*(1), 45–64.

Harvey, A. C. and N. Shephard (1996). Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business & Economic Statistics 14*(4), 429–34.

Jacquier, E., N. G. Polson, and P. Rossi (2004). Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors. *Journal of Econometrics 122*, 185–212.

Jensen, M. J. and J. M. Maheu (2010). Bayesian Semiparametric Stochastic Volatility Modeling. *Journal of Econometrics 157*, 306–316.

Jensen, M. J. and J. M. Maheu (2011). Semiparametric Asymmetric Stochastic Volatility. Technical report.

Kalli, M., S. G. Walker, and P. Damien (2011). Modelling the Conditional Distribution of Daily Stock Index Returns: An Alternative Bayesian Semiparametric Model. Technical report, University of Kent, Centre for Health Services Studies.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood Inference and Comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: 1. Density Estimates. *Annals Of Statistics 12*, 351–357.

Nakajima, J. and Y. Omori (2009). Leverage, Heavy-Tails and Correlated Jumps in Stochastic Volatility Models. *Computational Statistics & Data Analysis 53*, 2335–2353.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics 9*, 249–265.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics 140*, 425–449.

Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes - A Study of Daily Sugar Prices 1961-79. In O. D. Anderson (Ed.), *Time Series Analysis: Theory and Practice 1*, pp. 203–226. North-Holland.