



# CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Kalli, M., Damien, P. and Walker, S. (2013) Modelling the conditional distribution of daily stock index returns: an alternative Bayesian semiparametric model. *Journal of Business and Economic Statistics*, 31 (4). ISSN 0735-0015.

Link to official URL (if available):

<http://dx.doi.org/10.1080/07350015.2013.794142>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: [create.library@canterbury.ac.uk](mailto:create.library@canterbury.ac.uk)



# Modelling the conditional distribution of daily stock index returns: an alternative Bayesian semiparametric model

Maria Kalli, Stephen G. Walker, and Paul Damien \*

July 11, 2013

## Abstract

This paper introduces a new family of Bayesian semi-parametric models for the conditional distribution of daily stock index returns. The proposed models capture key stylized facts of such returns, namely heavy tails, asymmetry, volatility clustering, and the ‘leverage effect’. A Bayesian nonparametric prior is used to generate random density functions that are unimodal and asymmetric.

---

\*Maria Kalli, Business School, Canterbury Christ Church University, Canterbury, UK (email: maria.kalli@canterbury.ac.uk), Stephen G. Walker, School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK (email: S.G.Walker@kent.ac.uk), and Paul Damien, McCombs Business School, University of Texas at Austin, USA (email: Paul.Damien@mcombs.utexas.edu). We would like to thank the referees, the Associate Editor and Editor for their valuable comments.

Volatility is modelled parametrically. The new model is applied to the daily returns of the S&P 500, FTSE 100, and EUROSTOXX 50 indices and is compared to GARCH, Stochastic Volatility, and other Bayesian semi-parametric models.

**Keywords:** Stick-breaking processes; Infinite uniform mixture; Markov chain Monte Carlo; Slice sampling.

## 1 Introduction

Financial managers as well as investors would like to be in a position to accurately forecast asset return movements. The underlying distribution of these returns affects their decisions. Their predictions on asset price movements will be inaccurate, leading to bad investment decisions, if the distributional assumptions are not reflected in the financial data they are interested in.

The challenging task of modelling the conditional distribution of asset returns has been the subject of many studies. Some develop ARCH-GARCH type models, first introduced in Engle (1982) and Bollerslev (1986), while others build on the stochastic volatility (SV) model first introduced by Taylor (1982). For a comprehensive list of models, see Tsay (2005), Taylor (2005), and Bollerslev (2008). The difficulty lies in developing a model that captures the complicated features of the returns' distribution, often referred to as stylized facts, (see Cont (2001); Poon and Granger (2003)). In this paper we build a GARCH-type model for the asset returns' conditional distribution, which has a mode and a mean close to zero, asymmetry, heavy tails, volatility clustering and 'leverage effects'.

The deterministic nature and ease of estimation of GARCH type models makes them a popular choice. The basic set up of these models is:

$$y_t = h_t \epsilon_t \text{ for } t = 1, 2, \dots, n \quad (1)$$

where  $y_t$  is the log return,  $\epsilon_t$  is the innovation following some distribution  $F_\epsilon$  with mean zero and variance  $\sigma_\epsilon^2 = 1$ , and  $h_t^2$  is the volatility at time  $t$ , which is a function of previous volatilities and returns. The choice of  $F_\epsilon$  determines the conditional distribution of returns and impacts on distributions of future returns.

The normal distribution had been a popular choice for  $F_\epsilon$ . However, it has been shown that GARCH-type models with normal innovations fail to capture the heavy tails and slight asymmetry of the conditional distribution of returns. Alternative distributional choices include the Student-t distribution (Bollerslev, 1987), the generalised error distribution (Nelson, 1990), and a mixture of normal distributions (Bai et al., 2003). Although these alternatives account for the heavy-tailed behaviour of the distribution, they fail to capture skewness. To account for both aforementioned features, Gallant and Tauchen (1989) used Gram-Charlier expansions, while Hansen (1994) introduced the skewed Student-t distribution, and Theodossiou (1998) considered the Generalised t-distribution. Recently Chen et al. (2011) used the Asymmetric Laplace distribution. However, all these choices are constrained by the properties and parameters of the distributional family to which they belong to.

This paper describes a Bayesian nonparametric approach to modelling the conditional distribution of returns by nonparametrically estimating  $F_\epsilon$ . Bayesian nonparametric models place a prior on an infinite dimensional parameter space

and adapt their complexity to the data. A more appropriate term is infinite capacity models, emphasising the crucial property that they allow their complexity (i.e. the number of parameters) to grow as more data are observed; in contrast, finite-capacity models assume a fixed complexity. For a book length review of Bayesian nonparametric methods see Hjort et al. (2010).

The Dirichlet Process Mixture (DPM) model, an infinite mixture model, first introduced by Lo (1984) is the most popular Bayesian nonparametric model in financial econometric applications, (see Chib and Hamilton (2002), Hirano (2002), Jensen (2004), Griffin and Steel (2006), Leslie et al. (2007), Shahbaba (2009), and Jensen and Maheu (2010)). The DPM model uses the Dirichlet process (DP) (Ferguson, 1973) as a prior over the parameters of a normal distribution, with density  $k(\cdot)$ , facilitating the modelling of complex densities  $f(\cdot)$ . Under the DPM model  $f(\cdot) = \int k(\cdot|\boldsymbol{\theta})G(d\boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is the parameter vector, and  $G$  is the unknown random distribution drawn from a DP.

In our approach to modelling the innovations' distribution,  $F_\epsilon$ , we depart from the DPM model in two ways. We use a stick-breaking prior (SBP) instead of a DP to generate  $G$ , and substitute the normal density with a scaled uniform density. The uniform density ensures unimodality and an asymmetry parameter allows us to capture the asymmetry in asset returns. Stick-breaking processes are more general than the DP, in fact the DP is a subclass. They can therefore adapt more flexibly to the data and together with the scaled uniform density can capture more accurately the heavy tailed behaviour. We believe that with this model we can better account for extreme events, which impact on the tail behaviour of the returns' distribution, by mitigating the risk of placing artificial modes at unusual points. We model the volatility dynamics using: the

GARCH(1,1), the GJR-GARCH(1,1) and the EGARCH(1,1) models; the latter two can capture the ‘leverage effect’.

The structure of this paper is as follows, in Section 2 we describe in detail our infinite uniform mixture (IUM) model based on stick-breaking priors for GARCH-type volatility representations. Section 3 describes the sampling methodology for the IUM model, Section 4 provides a simulation study comparing IUM to a SBP with a normal set up and to the DPM. In Section 5 we analyse samples from the S&P 500, FTSE 100, and EUROSTOXX 50 daily index returns using our IUM model and provide comparisons with GARCH, SV models and alternative Bayesian semiparametric models. We summarise our conclusions in Section 6.

## 2 GARCH-type infinite uniform mixture model (IUM)

Bayesian infinite mixture models were popularized by Lo (1984). The problem of estimating a density  $f(\cdot)$  is addressed using a Bayesian nonparametric prior over the parameters of a continuous density function  $k(\cdot)$ . The model is represented by

$$f(\cdot) = \int k(\cdot|\boldsymbol{\theta})G(d\boldsymbol{\theta}),$$

where  $\boldsymbol{\theta}$  is the parameter vector, and  $G$  is an unknown random distribution drawn from a Bayesian nonparametric prior.

Stick-breaking priors are examples of Bayesian nonparametric priors. They are discrete random probability measures represented by

$$G(\cdot) = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}(\cdot), \quad (2)$$

where  $\delta_{\theta_j}$  is the Dirac measure giving mass one at location  $\theta_j$ , with weight  $w_j$ . The weights must satisfy two conditions in order for  $G$  to be a probability measure:  $0 < w_j < 1$  and  $\sum_{j=1}^{\infty} w_j = 1$  with probability one. The locations  $\theta_j$  are independent identically distributed (*iid*) random variables with distribution  $G_0$  and are independent of the random weights  $w_j$ . The distribution  $G_0$  is often referred to as the base or centering distribution. This is so because for any measurable set  $B$  of a  $\sigma$ -field  $\mathcal{B}$  we have  $E[G(B)] = G_0(B)$ . The random weights  $w_j$  are transformations of *iid* random variables,  $v_j \sim \text{Be}(a_j, b_j)$ , that is

$$w_1 = v_1 \text{ and } w_j = (1 - v_1) \dots (1 - v_{j-1}) v_j = v_j \prod_{l < j} (1 - v_l). \quad (3)$$

Because of their representation they are referred to as stick-breaking weights. For more on the long history of the stick-breaking notion of constructing infinite dimensional priors see Halmos (1944), Freedman (1963), Kingman (1974), and Ishwaran and James (2001).

The Dirichlet process is a subclass of stick-breaking processes. It arises when  $a_j = 1$  and  $b_j = c$ , where  $c \in [0, \infty)$ . The parameter  $c$  is often referred to as the DP concentration parameter because it controls how close  $G$  is to  $G_0$ . It also controls the rate of decay of the weights. Looking at the expectation of the weights we have

$$E(w_j) = \frac{1}{1+c} \left(1 - \frac{1}{1+c}\right)^{j-1} \text{ for } j > l,$$

and we can clearly see that the value of  $c$  solely controls their decay. This decay is exponential, leading to fewer mixture components with non-negligible

weights as  $j$  increases. This can be a disadvantage as more mixture components may be needed to capture the heavy tails of the conditional returns' distribution. Stick-breaking priors are more flexible. The number of non-negligible  $w_j$  depends on two parameters, the beta parameters  $(a_j, b_j)$ , rather than one. These parameters also depend on  $j$ , and therefore we have an infinite number of parameters controlling the rate of decay, which provides more flexibility to the model as more mixture components with small weights could be used to account for the heavy tail behaviour of the returns' distribution. This is the reason we have decided to use a SBP as the generating process for  $F_\epsilon$ . In our illustrations we specify the parameters of the parameters of this prior in the following way. If  $w_j \sim SBP(a_j, b_j)$  then

$$E[w_j] = \frac{a_j}{a_j + b_j} \prod_{l < j} \left( 1 - \frac{a_l}{a_l + b_l} \right) \text{ for } j > l.$$

We will center this process over a distribution for the weight by choosing  $E[w_j] = \xi_j$ , where  $\xi_j$  is  $Pr(X = j)$  for a random variable  $X$  with a discrete distribution on  $1, 2, 3, \dots$ . The random variable  $X$  is given a Beta-Geometric distribution i.e.  $Pr(X = j) = p(1 - p)^{j-1}$  where  $p \sim Be(\alpha_1, \alpha_2)$ . This yields

$$\xi_j = \frac{\Gamma(\alpha_1 + \alpha_2)\Gamma(\alpha_1 + 1)\Gamma(\alpha_2 + j - 1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_1 + \alpha_2 + j)},$$

which allows us to control the number of non-negligible weights by choosing the values of  $(\alpha_1, \alpha_2)$ . Appendix 1 provides more details on this specification.

The other novel contribution of this paper is our choice of continuous density function  $k(\cdot)$ , and centering distribution  $G_0$ . For the DPM model, both  $k(\cdot)$ , and  $G_0$  are the density of a normal distribution and a normal distribution respectively. We propose an infinite mixture of uniforms for the density of the



innovations,  $f_\epsilon(\cdot)$  represented by

$$f_\epsilon(\cdot) = \int \nu(\epsilon|\theta, \lambda)G(d\theta), \quad (4)$$

where  $\nu(\epsilon|\theta, \lambda)$  is the density function of the scaled uniform distribution  $U(-e^{-\lambda}\theta, e^\lambda\theta)$  with asymmetry parameter  $\lambda$  and location  $\theta$ . The unknown distribution  $G$  is generated from a  $SBP(a_j, b_j)$  with a standard exponential distribution as the centering distribution  $G_0$ . With the representation of equation (4) we ensure unimodality of the conditional return distribution, model extreme returns using heavy tails and avoid the risk of artificial modes at these extreme returns as it may be the case under the DPM model. Take the simplest case of  $U(-\theta, \theta)$ . This ensures unimodality for the innovations' distribution with mode at zero. The random distribution  $G$  ranges over all distribution functions on  $(0, \infty)$  and therefore  $f_\epsilon$  ranges over all unimodal and symmetric density functions on  $(-\infty, \infty)$ , see Feller (1957). We account for the slight asymmetry of returns with parameter  $\lambda$  following Fernandez and Steel (1998). We can have both negative and positive skewness, and symmetry when  $\lambda < 0$ ,  $\lambda > 0$ , and  $\lambda = 0$  respectively. The flexible construction of our infinite mixture of uniforms means that we can develop a unimodal model family which captures any level of kurtosis, and accounts for the slight asymmetry in the returns  $y_t$ .

Under our stick breaking representation, the infinite mixture of uniforms for modelling the distribution of the innovations has the following hierarchical set up:

$$y_t = h_t \epsilon_t$$

where  $y_t$  are the log returns

$$\epsilon_t \sim U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^\lambda) \text{ for } t = 1, \dots, n \quad (5)$$

$$\Pr(d_t = j) = w_j \text{ for } j = 1, 2, \dots$$

$\theta_j \sim G_0(\cdot)$  for  $j = 1, 2, \dots$  where  $G_0$  is a standard exponential distribution

$$w_1 = v_1 \text{ and } w_j = (1 - v_1) \dots (1 - v_{j-1})v_j = v_j \prod_{l < j} (1 - v_l)$$

$$v_j \sim \text{Be}(a_j, b_j).$$

Thus the distribution of  $\epsilon_t$  is

$$f_{v,\theta}(\epsilon_t) = \sum_{j=1}^{\infty} w_j U(-\theta_j e^{-\lambda}, \theta_j e^{\lambda}), \quad (6)$$

with mean

$$\frac{(e^{\lambda} - e^{-\lambda}) \sum_{j=1}^{\infty} w_j \theta_j}{2}$$

and variance

$$\frac{4(e^{2\lambda} + e^{-2\lambda} - 1) \sum_{j=1}^{\infty} w_j \theta_j^2 - 3(e^{2\lambda} + e^{-2\lambda} - 2) (\sum_{j=1}^{\infty} w_j \theta_j)^2}{12}.$$

This implies that the conditional density of log return  $y_t$  is represented by

$$f_{G,\lambda}(y_t | h_t) = \sum_{j=1}^{\infty} w_j U(y_t | -\theta_j h_t e^{-\lambda}, \theta_j h_t e^{\lambda}).$$

For our illustrations we will use three GARCH-type models for volatility,  $h_t$ :

1. The GARCH(1,1) of Bollerslev (1986) where

$$h_t^2 = \beta_0 + \beta_1 y_{t-1}^2 + \phi h_{t-1}^2. \quad (7)$$

2. The GJR-GARCH(1,1) of Glosten et al. (1993) where

$$h_t^2 = \beta_0 + \beta_1 y_{t-1}^2 + \beta_2 I_{t-1} y_{t-1}^2 + \phi h_{t-1}^2 \quad (8)$$

with  $I_{t-1} = 1$  for  $y_{t-1} < 0$  and  $I_{t-1} = 0$  for  $y_{t-1} \geq 0$ .

3. The EGARCH(1,1) of Nelson (1990) where

$$\log(h_t^2) = \beta_0 + \beta_1 \left( \frac{|y_{t-1}|}{h_{t-1}} - E \left[ \frac{|y_{t-1}|}{h_{t-1}} \right] \right) + \beta_2 \frac{y_{t-1}}{h_{t-1}} + \phi \log(h_{t-1}^2). \quad (9)$$

All models are characterised by an ARCH parameter  $\beta_1$  and a volatility parameter  $\phi$ . The sum of  $\beta_1$  and  $\phi$  can be interpreted as a measure of persistence of shocks to volatility. The latter two representations have an extra parameter  $\beta_2$  which is introduced to capture the asymmetric response of volatility to positive and negative shocks to returns i.e. the ‘leverage effect’. In the case of GJR-GARCH(1,1) a  $\beta_2 > 0$  signifies the capture of the ‘leverage effect’, and in the case of EGARCH(1,1) a  $\beta_2 < 0$ . From now on we will refer to  $\beta_0, \beta_1, \beta_2$  and  $\phi$  as volatility coefficients. To ensure stationarity for  $y_t$ , conditions are imposed on these. For GARCH(1,1)  $\beta_0 > 0, \beta_1 > 0, \phi > 0$  and  $\beta_1 + \phi < 1$ , for GJR-GARCH  $\beta_0 \geq 0, \beta_1 \geq 0, \phi \geq 0, \beta_1 + \beta_2 \geq 0$ , and  $\beta_1 + \frac{\beta_2}{2} + \phi < 1$ , and for EGARCH(1,1) we need to ensure that  $|\phi| < 1$ . The additional requirement for EGARCH(1,1) is the calculation of  $E \left[ \frac{|y_{t-1}|}{h_{t-1}} \right]$ . For this paper we base this calculation on the prior distribution of  $\theta$ , and equation (6), that is  $\sum_{j=1}^{\infty} w_j U(-\theta_j e^{-\lambda}, \theta_j e^{\lambda})$ . The resulting formula is

$$E \left[ \frac{|y_{t-1}|}{h_{t-1}} \right] = \frac{e^{2\lambda} + e^{-2\lambda}}{2(e^{\lambda} + e^{-\lambda})} \sum_{j=1}^{\infty} w_j \theta_j, \quad (10)$$

where  $\sum_{j=1}^{\infty} w_j \theta_j$  reduces to 1 as the prior for  $\theta_j$  is the standard exponential distribution.

### 3 Computation

This section details the MCMC algorithm for fitting the GARCH-type IUM model. Under our model the mixing distribution  $F_{\epsilon}$  is modelled by a stick-breaking

prior, leading to a stick-breaking uniform mixture distribution for the innovations,  $\epsilon_t$ . The infinite number of mixture components together with the fact that our likelihood and the prior do not form a conjugate pair (as is the case with the DPM) make inference difficult. Our MCMC algorithm provides samples from the joint posterior distribution of the model parameters, and makes use of the slice-efficient sampler of Kalli et al. (2011) to address the issue of infinite mixture components and weights of the hierarchical model (5) and equation (6). The exact number of components and weights that we need in order to produce a valid Markov chain with the correct stationary distribution is found by introducing a latent variable  $u_t$  (to ensure that the number of mixture components is finite), a decreasing positive sequence  $(\zeta_j)$  (to address the updating of the  $u_t$ ), and an allocation variable  $d_t$  which indicates which of the mixture components have innovations allocated to them. For more details on the slice-efficient sampler and its mixing performance see Kalli et al. (2011). Following the introduction of  $u_t$ ,  $\zeta_t$ , and  $d_t$ , the joint density of  $\epsilon_t$ ,  $u_t$ , and  $d_t$  is then

$$f_{v,\theta}(\epsilon_t, u_t, d_t) = \zeta_{d_t}^{-1} \mathbf{1}(u_t < \zeta_{d_t}) w_{d_t} U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^{\lambda}). \quad (11)$$

The variables that need to be sampled at each iteration of this Gibbs algorithm are

$$\{(\theta_j, v_j), j = 1, 2, \dots; (d_t, u_t), t = 1, \dots, n\},$$

and the joint posterior distribution is then proportional to:

$$\prod_{t=1}^n \mathbf{1}(u_t < \zeta_{d_t}) w_{d_t} \zeta_{d_t}^{-1} U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^{\lambda}).$$

Since  $\zeta$  and  $v$  are conditionally independent, the sampling steps are:

- $\pi(\theta_j | \dots) \propto g_0(\theta_j) \prod_{d_t=j} U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^{\lambda})$

- $\pi(v_j) \propto \text{beta}(v_j; a'_j, b'_j)$ , where  $a'_j = a_j + \sum_{t=1}^n \mathbf{1}(d_t = j)$  and  $b'_j = b_j + \sum_{t=1}^n \mathbf{1}(d_t > j)$ .
- $\pi(u_t | \dots) \propto \mathbf{1}(0 < u_t < \zeta_{d_t})$ .
- $\Pr(d_t = \kappa | \dots) \propto \mathbf{1}(\kappa : \zeta_\kappa > u_t) w_\kappa \zeta_\kappa^{-1} U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^\lambda)$

We complete this section by detailing the sampling steps for  $\theta, \lambda$  and the volatility coefficients. For simplicity we will explain the sampling scheme for the volatility coefficients when we have the GARCH(1,1) set up where

$$h_t^2 = b_0 + b_1 y_{t-1}^2 + \phi h_{t-1}^2,$$

noting that for the GJR-GARCH (1,1) and EGARCH(1,1) the sampling is the same but with the additional parameter  $b_2$  which accounts for the ‘leverage effect’.

Recall that  $y_t = h_t \epsilon_t$  and thus the joint posterior distribution we will sample from is proportional to

$$\prod_{t=1}^n \mathbf{1}(u_t < \zeta_{d_t}) w_{d_t} \zeta_{d_t}^{-1} U(-\theta_{d_t} h_t e^{-\lambda}, \theta_{d_t} h_t e^\lambda).$$

### Updating the $\theta$ 's

The  $\theta$ 's are generated from  $G_0$  which is a standard exponential distribution, thus their full conditional is proportional to

$$e^{-\theta_j} \frac{\prod_{d_t=j} \mathbf{1}\{-\theta_j e^{-\lambda} h_t < y_t < \theta_j e^\lambda h_t\}}{\theta_j^{n_j}} \quad (12)$$

where  $n_j$  is the number of  $y_t$  allocated to cluster  $j$ , which is the size of the cluster formed by those  $d_t = j$ . The tricky part is to correctly identify the truncations imposed by  $\mathbf{1}\{-\theta_j e^{-\lambda} h_t < y_t < \theta_j e^\lambda h_t\}$ , by considering the case when  $y_t$  is negative and the case when  $y_t$  is positive. So the truncation when

$y_t$  is negative is  $\theta_j > \max_{d_t=j, y_t < 0} \left\{ \frac{-y_t}{e^{-\lambda} h_t} \right\}$  and when  $y_t$  is positive it is  $\theta_j > \max_{d_t=j, y_t > 0} \left\{ \frac{y_t}{e^{\lambda} h_t} \right\}$ . We can therefore define the truncation point for  $\theta_j$  as

$$tr_{\theta_j} = \max_{d_t=j} \left\{ \max_{y_t > 0} \left( \frac{y_t}{e^{\lambda} h_t} \right), \max_{y_t < 0} \left( \frac{-y_t}{e^{-\lambda} h_t} \right) \right\}. \quad (13)$$

A rejection sampler is used to update the  $\theta_{d_t}$ 's from (12). The candidate density is a truncated exponential with  $tr_{\theta_j}$  being the truncation. For the case where  $j > d_t$  we sample the  $\theta$ 's from the prior, which is the standard exponential.

### Updating $\lambda$

For the updates of the skewness parameter  $\lambda$  we again need to consider the truncations that arise due our choice of a uniform kernel. We take a normal prior with mean 0 and variance  $1/2s$  to capture the slight asymmetry of returns. In our illustrations  $s = 2$ . The full conditional for  $\lambda$  is therefore proportional to

$$\frac{e^{-s\lambda^2}}{(e^{-\lambda} + e^{\lambda})^n} \prod_{t=1}^n \mathbf{1}\{-\theta_{d_t} e^{-\lambda} h_t < y_t < \theta_{d_t} e^{\lambda} h_t\}. \quad (14)$$

To define the truncations we again consider the case when  $y_t$  is negative and the case when  $y_t$  is positive. For a negative  $y_t$  the truncation is

$$tr_{-\lambda} = \min_{y_t < 0} \{-\log(y_t) + \log(-\theta_{d_t}) + \log(h_t)\}.$$

For a positive  $y_t$  the truncation is

$$tr_{\lambda} = \max_{y_t > 0} \{\log(y_t) - \log(\theta_{d_t}) - \log(h_t)\}.$$

Since (14) is a log concave function we use the Adaptive Rejection Sampler (ARS) of Gilks and Wild (1992) to update  $\lambda$  subject to the truncation points  $tr_{\lambda}$  and  $tr_{-\lambda}$ .

## Updating the volatility coefficients

The joint conditional for the volatility coefficients is proportional to

$$\prod_{t=1}^n \mathbf{1} \left\{ -\theta_{d_t} e^{-\lambda} h_t < y_t < \theta_{d_t} e^{\lambda} h_t \right\} \frac{1}{h_t}. \quad (15)$$

We draw attention to the fact that not all clusters have innovations allocated to them. We have empty and full clusters, the full being identified by the  $\theta_{d_t}$ . For updating the volatility coefficients we consider the full clusters. To improve the mixing of the Markov chain we propose integrating over the  $\theta_{d_t}$  and working with a simpler likelihood. Our starting point is

$$\pi(y|h) \propto \prod_{j=1}^{N_c} \frac{1}{\prod_{t=1}^n h_t} \int \frac{1}{\theta_j^{n_j}} \mathbf{1} \{ \theta_j > tr_{\theta} \} e^{-\theta_j} d\theta_j, \quad (16)$$

where  $N_c$  is the number of full clusters. As with the update of the  $\theta$ 's our prior is the standard exponential,  $n_j$  is the number of  $y_t$  allocated to cluster  $j$  and  $tr_{\theta_j}$  is expressed in equation (13). In order to simplify (16) we introduce auxiliary variables,  $z_j$ . The distribution of  $z_j$  conditional on the  $\theta_j$  is gamma, that is  $z_j|\theta_j \sim \text{Ga}(n_j, \theta_j)$ . The likelihood then becomes

$$\pi(y|h) \propto \frac{1}{\prod_{t=1}^n h_t} \prod_{j=1}^{N_c} \int_{tr_{\theta_j}}^{\infty} z_j^{n_j-1} e^{-\theta_j(z_j+1)} d\theta_j.$$

Integrating over the  $\theta_j$ , the simpler likelihood which we use to update the volatility coefficients is

$$\pi(y|h) \propto \frac{1}{\prod_{t=1}^n h_t} \prod_{j=1}^{N_c} z_j^{n_j-1} \frac{e^{-tr_{\theta_j}(z_j+1)}}{z_j + 1}. \quad (17)$$

We use the random walk Metropolis Hastings (MH) sampler to update each of the volatility coefficients. In the case when  $h_t^2$  is represented as a GARCH(1,1) and GJR-GARCH(1,1) we incorporate their stationarity conditions in the sampler. For GARCH(1,1) we ensure that  $\beta_0 > 0$ ,  $\beta_1 > 0$ ,  $\phi > 0$  and  $\beta_1 + \phi < 1$ ,

and for GJR-GARCH(1,1) we ensure that  $\beta_0 \geq 0$ ,  $\beta_1 \geq 0$ ,  $\phi \geq 0$  and  $\beta_1 + \beta_2 \geq 0$  and  $\beta_1 + \frac{\beta_2}{2} + \phi < 1$ . We choose a normal distribution truncated to interval  $(0, 1)$  as the candidate density  $q(\cdot)$  for each of these coefficients, with equation (17) as the target density. We then accept the new values  $\beta'_0, \beta'_1$ , and  $\phi'$  with probability

$$\alpha(\phi, \phi') = \frac{\pi(y|h')q(\phi', \phi)}{\pi(y|h)q(\phi, \phi')}.$$

We used  $\phi$  to show the acceptance ratio, as the set up of it is the same for all coefficients, just replace  $\phi$  with  $\beta_1$  and  $\beta_0$ . In the case where  $h_t^2$  has the EGARCH(1,1) representation we ensure that  $|\phi| < 1$  and use equation (9) with a multivariate normal as the candidate density  $q(\cdot)$ .

Updating these coefficients one at a time is a lengthy process due to the slow mixing of the MH sampler. This is because the correlation between the coefficients is high, and mixing is improved by sampling the volatility coefficients in a block (see for example Chen and So (2006)). To efficiently sample from the block we use the ideas of adaptive Monte Carlo discussed in Roberts and Rosenthal (2009). Adaptive Monte Carlo algorithms allow the proposal distribution in a Metropolis-Hastings step to change over an MCMC run. This allows the sampler to adapt to the form of the posterior. The sampler is no longer Markovian and so care is needed to define an algorithm that converges to the posterior. Roberts and Rosenthal (2009) develop a general theory for convergence of these algorithms using two conditions: diminishing adaptation and uniform ergodicity. We do not discuss these concepts further since the methods used in this paper are well studied and convergence is proved in the accompanying references.



Let  $\varphi = (\beta_0, \beta_1, \phi)$  which allows us to use the random walk MH with a multivariate normal as the candidate density as described in Haario et al. (2001). Our multivariate normal distribution has mean  $\varphi$  and covariance matrix  $c_\varphi \Sigma_{it}$ . The positive constant  $c_\varphi$  is introduced to secure good acceptance rates. Though Haario et al. (2001) suggest  $c_\varphi = 2.34$  they allow the modeller to choose its value based on trial and error. In our illustrations we used  $c_\varphi = 1$ . The covariance matrix  $\Sigma_{it}$  is the posterior covariance matrix at each iteration, that is

$$\Sigma_{it} = \frac{1}{it-2} \left\{ \sum_{i=1}^{it-1} \varphi^{(i)} \varphi^{(i)T} - \frac{(\sum_{i=1}^{it-1} \varphi^{(i)})(\sum_{i=1}^{it-1} \varphi^{(i)})^T}{it-1} \right\}. \quad (18)$$

where  $it = 1, \dots, M$  is the iteration index. This set up allows the covariance matrix to adapt with each update of the volatility coefficients, and thus provide a candidate density at each iteration,  $q_{it}(\varphi, \varphi')$ , that adapts with each update of the coefficients and of the covariance matrix. This results in improved mixing of the MCMC sampler.

This blocked update for  $\varphi$  uses estimates of the variances and covariances for all volatility coefficients. We suggest using an alternative update for a few iterations (in our examples 200 iterations) in order to allow these estimates to settle. Using the idea of Atchade and Rosenthal (2005) we adjust the variances of the candidate densities of each of the volatility coefficients as follows, (we display the set up using  $\phi$  but this is exactly the same for  $\beta_0$  and  $\beta_1$ )

$$\sigma_\phi^2 = \sigma_\phi^2 + \frac{1}{\sqrt{it}}(\alpha - 0.3),$$

where  $\alpha$  is the acceptance ratio and  $it$  is the index of the iterations,  $it = 1, \dots, M$ . This correction allows the variance to adapt with each acceptance step at each iteration and thus improve on the mixing. The choice of 0.3 is a

conservative one and it ensures that  $\alpha$  does not fall below 0.3. Roberts et al. (1997) find that the optimal acceptance rate for random walk MH samplers is 0.234, however in Roberts and Rosenthal (2009) it is suggested that it is best in the case of the aforementioned adaptive step to be conservative.

## 4 Simulation Example

To assess the fit and predictive performance of the stick-breaking prior uniform mixture, hereafter referred to as IUM, we simulated a single time series for  $y_t$  from the model in equation (1), where  $t = 1, \dots, 3000$ . Volatility was simulated by the GARCH(1,1) model

$$h_t^2 = 0.01 + 0.15y_{t-1}^2 + 0.80h_{t-1}^2, \quad (19)$$

and the innovations were generated from the mixture distribution

$$\epsilon_t \sim 0.9N(0.1, 0.5) + 0.1N(-1, 4.41).$$

We compared the IUM model to the DPM model and to a stick-breaking prior (SBP) with a normal mixture set up. To facilitate inference for both of these alternatives we adjusted our computation to account for the normal mixture set up. We provide the details of this adjustment in Appendix 2. We ran the MCMC sampler under all three set ups using 35000 iterations and discarded the initial 5000 as burn in. Trace plots (Figure 1) and the median estimates together with the 95% credible intervals of the posterior samples of the volatility parameters,  $\beta_0$ ,  $\beta_1$ , and,  $\phi$  (Table 2) were generated.

The trace plots show the mixing of the volatility coefficients under the three Bayesian nonparametric models. It is very similar for the DPM and SBP-normal, and for the IUM it is also good. The volatility estimates of the IUM model are the ones that are closer to the true values. Though the other two models, DPM

and SBP-normal do well in estimating the value of  $\phi$ , they underestimate the values of  $\beta_0$  and  $\beta_1$ . Finally the 95% credible intervals under the DPM and the SBP-normal have similar width whereas those of the IUM model are actually narrower.

To compare the fit of each of the three competing models, we calculated the Mean Integrated Squared Error (MISE). Classically the  $MISE = \int (f_\epsilon^{true}(x) - \hat{f}_\epsilon(x))^2 dx$  where  $f_\epsilon^{true}(x)$  is the true density of the innovations and  $\hat{f}_\epsilon(x)$  the estimated density. The model with the smallest MISE is preferred. In our Bayesian approach MISE is viewed as the posterior expectation of the distance between the true density estimated at point  $x$  and  $f_\epsilon$  evaluated at  $x$ , that is  $E[\int (f_\epsilon^{true}(x) - f_\epsilon(x))^2 dx|y]$ . We approximate this expectation by

$$\frac{1}{M} \sum_{it=1}^M \left[ \int_{lo}^{up} (f_\epsilon^{true}(x) - f_\epsilon^{(it)}(x))^2 dx \right],$$

where  $it = 1, \dots, M$  is the index of the iterations, and  $f_\epsilon^{(it)}$  is the estimated density at iteration  $it$ . We ensure that  $F_\epsilon^{true}(lo) = 10^{-5}$  and  $F_\epsilon^{true}(up) = 1 - 10^{-5}$ , and evaluate the integral using the trapezoidal rule. The MISE estimates are also displayed in Table 2. From the comparison of the three MISE estimates it is clear that the IUM has the best fit. These results for both the MISE and the volatility estimates demonstrate the flexibility of the IUM over the two normal set ups, the DPM and the SBM-normal. Having a uniform mixture coupled with the ability to control the decay of the weights leads to better fitting models and volatility estimates that are closer to the true values of the model from where the data was generated.

We assessed the predictive performance of the competing models by calculating both average log predictive scores (LPS) and average log predictive

tail scores (LPTS) (Delatola and Griffin, 2011). These scores are based on the one step ahead predictive density  $f(y_{t+1}|y_{(1:t-1)}, \vartheta)$ , where  $\vartheta$  represents the model parameters;  $\varphi$  the vector of the volatility coefficients, and  $F_\epsilon$ , the innovation distribution. We approximate the predictive density by  $f(y_{t+1}|y_{(1:t-1)}, \hat{\vartheta})$ , where  $\hat{\vartheta}$  represents parameter estimates. For  $\varphi$  these are the posterior medians of each volatility coefficient. For  $F_\epsilon$  the point estimate is  $\widehat{F_\epsilon(B)} = E[F_\epsilon(B)]$  for measurable sets  $B$ , which is the Bayes estimator under squared loss.

For the calculation of LPS and LPTS  $t = 1, \dots, n$  refers to the latter half of the data set, the evaluation (out of sample) set. The first half is the training (in sample) set which is used to estimate  $\varphi$ , and  $F_\epsilon$ . We calculate LPS as

$$\text{LPS} = -\frac{1}{n} \sum_{t=1}^n \log f(y_t | y_{1:(t-1)}, \hat{\vartheta}). \quad (20)$$

Under this calculation smaller values of the LPS indicate a better fitting model. In practice we may be more interested in predicting extreme returns and thus we calculate the LPTS. We identify  $z_\alpha$ , the upper 100 $\alpha$ % of the absolute values of the standardised  $y_t$ . Since we built a conditional model and are interested in the conditional ‘tails’, we carried out this standardisation, in order to have a measure of the conditional extreme returns. The standardised  $y_t$  is calculated as  $y_t^* = \frac{y_t}{\hat{h}_t}$ , where  $\hat{h}_t$  is an estimate of volatility  $h_t$ , calculated using the estimates of the volatility coefficients from the evaluation set. We then find  $\mathbf{1}(|y_t^*| \geq z_\alpha)$ , the number of extreme returns of  $y_t^*$ . Thus only predictions of returns with absolute value in the upper 100 $\alpha$ % of  $|y_t^*|$  are included in the score. So, LPTS is,

$$\text{LPTS} = -\frac{1}{\sum_{t=1}^n \mathbf{1}(|y_t^*| > z_\alpha)} \sum_{t=1}^n \mathbf{1}(|y_t^*| > z_\alpha) \log f(y_t | y_{1:(t-1)}, \hat{\vartheta}) \quad (21)$$

	LPS	LPTS-01	LPTS-05
DPM	0.311	3.435	4.700
SBP-normal	0.320	3.437	4.710
SBP-uniforms (IUM)	0.297	3.077	4.134

Table 1: LPS and LPTS under the three models for the simulated data.

	$\beta_0$	$\beta_1$	$\phi$	MISE
<i>True value</i>	<i>0.010</i>	<i>0.150</i>	<i>0.800</i>	
DPM	0.007	0.105	0.776	0.015
	(0.002,0.017)	(0.023,0.226)	(0.718,0.823)	(0.0003,0.1258)
SBP-normal	0.009	0.100	0.777	0.016
	(0.003,0.019)	(0.042,0.245)	(0.727,0.808)	(0.0004,0.1176)
IUM	0.011	0.165	0.797	0.007
	(0.004,0.015)	(0.085,0.219)	(0.758,0.824)	(0.0043,0.0234)

Table 2: Median estimates of volatility coefficients and MISE with their 95% credible intervals.

In our comparisons we looked at the upper 5% and 1% values. The results are displayed in Table 1. The IUM outperforms the other two models both in terms of the LPS and the LPTS. Its' lower LPTS values demonstrate that it can predict extreme returns more accurately than the two alternatives.

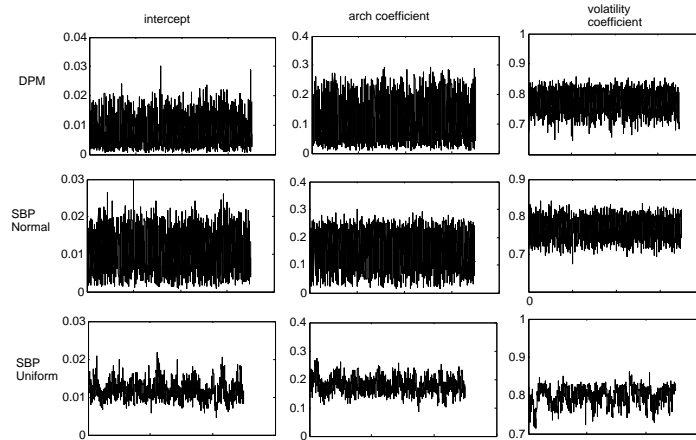


Figure 1: Trace plots of the posterior samples of the volatility coefficients,  $\beta_0, \beta_1$  (arch coefficient) and  $\phi$  (volatility coefficient).

## 5 Empirical Examples

In this paper we look at daily log-returns of three stock indices: the S&P 500, FTSE 100, and EUROSTOXX 50. Our samples for each index respectively are from: January 3<sup>rd</sup> 1980 to December 30<sup>th</sup> 1987, January 3<sup>rd</sup> 1997 to March 12<sup>th</sup> 2009, and June 7<sup>th</sup> 2002 to March 3<sup>rd</sup> 2009. Figure 2 shows the time plots for the daily log-returns. Table 3 shows the main summary statistics of these returns.

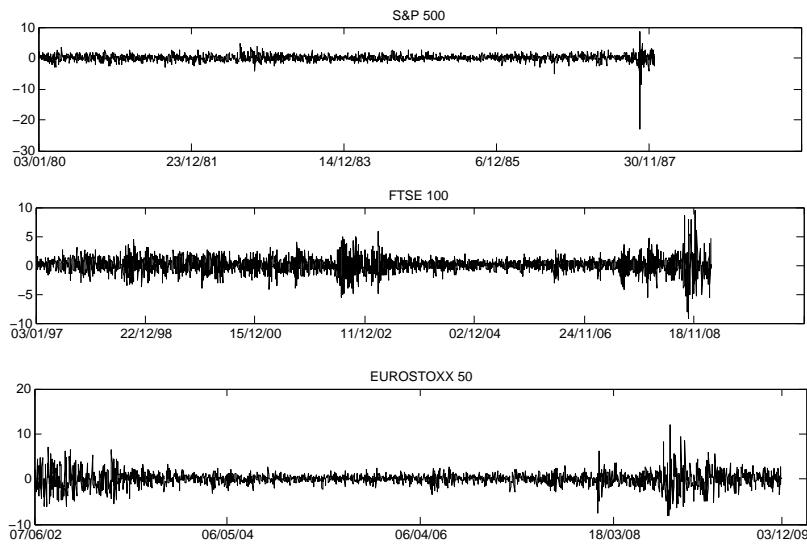


Figure 2: Time plots of observed daily returns

Descriptives	S&P 500	FTSE 100	EUROSTOXX 50
median	0.044	0.031	0.000
st.dev	1.130	1.304	1.610
skewness	-4.129	-0.124	0.010
kurtosis	90.609 (10.147)	8.772	8.929

Table 3: Summary Statistics

The plots in Figure 2 identify periods of sustained high volatility, which we

refer to as volatility clusters. The most notable volatility clusters for each index are those near the end of the sampled time period, which are all related to stock market crashes. For the S&P 500, it was the October 1987 crash whereas for the FTSE 100 and EUROSTOXX 50, it was the 2008 crash caused by the collapse of the subprime mortgage market. From the summary statistics displayed in Table 3 we focus on the estimates of kurtosis and skewness. All three samples have high levels of kurtosis. For the S&P 500 estimated kurtosis is nine times higher when the extreme drop in the prices on October 19<sup>th</sup> 1987 is in the sample than when it is excluded. This extreme observation also impacts on skewness, which increases to  $-0.022$  from  $-4.129$  once the extreme is removed. The estimated skewness from the EUROSTOXX 50 sample is positive  $0.010$  rather than negative as it the case with the samples from the other two indices.

We apply our infinite uniform mixture (IUM) with the GARCH(1,1), GJR-GARCH(1,1), and EGARCH(1,1) representations for the volatility to all three samples. We then compare our results to those from GARCH(1,1), GJR-GARCH(1,1) and EGARCH(1,1) with Student-t innovations, and skewed Student-t innovations. We refer to these models as parametric models. They were fitted using Matlab's Econometrics and Kevin Shephard's Financial Econometrics toolboxes. Both of these toolboxes carry out maximum likelihood estimation (MLE). Further comparisons were made with two SV models with Student-t innovations one without (SV(1)-t), and one with 'leverage effect' (SV(1)-t leverage), using the algorithms of Jacquier et al. (1994, 2004). Finally, we compared our IUM model with a DPM model with the same GARCH-type representations for the volatility, the SV-DPM model of Jensen and Maheu (2010), and the  $\pi$ DDP model of Griffin and Steel (2006). In all cases the priors and samplers sug-

gested by the authors are used. We refer to these models as Bayesian non-parametric models. The results for the IUM and DPM models were based on 480000 runs with the first 20000 discarded as burn in. The parameters  $a_j, b_j$  of the SBP are generated from the geometric-beta model (described in Section 2) with parameters  $\alpha_1 = 1, \alpha_2 = 6$ . For the DPM, the ‘precision’ parameter,  $c = 1$ .

## 5.1 Results

As was the case with the simulated example, we assess the predictive performance of the competing models by calculating both average log predictive scores (LPS) and log predictive tail scores (LPTS) as per equations 20 and 21. For the LPTS, we again considered the 5% and 1% points. The volatility estimates used to obtain  $y^*$  (the standardised  $y_t$ ) for each index were based on the GJR-GARCH(1,1) model with skewed student-t innovations. We did estimate the volatility using all GARCH-type models, and since the estimates were similar we decided to use those of the GJR-GARCH(1,1) model. The LPS and LPTS scores for each model and each data set are included in Table 4.

Looking at these scores we see that the IUM model for the innovation distribution with the three GARCH-type volatility set ups is a competitive model. It produces the lowest (best) LPS for two of the indices studied, the FTSE 100 and Eurostoxx 50 and the second best LPS for the S&P 500. In terms of the LPTS it outperforms all other models for the three indices. Based on the LPTS both the GJR-GARCH(1,1)-IUM and GARCH(1,1)-IUM have better predictive performance of extreme events when compared to the EGARCH(1,1)-IUM. All the GARCH-type IUM models produce better LPS and LPTS compared to the



respective GARCH-type models with DPM innovations and SV-DPM for all indices.

In Table 5 we also provide in-sample estimates of the volatility coefficients for all three sampled indices. For the point estimates of the volatility coefficients under the Bayesian models we calculate the posterior medians and also provide their 95% credible intervals. The  $\pi$ DDP of Griffin and Steel (2006) models the volatility nonparametrically and therefore there are no volatility coefficients to estimate. For the parametric models we calculate these estimates based on MLE and provide their 95% confidence intervals. Regarding the SV(1)-t with ‘leverage’ we do not include a value for a leverage coefficient. The ‘leverage effect’ in this model is estimated by the correlation between the innovations and the noise-terms of the volatility. We do provide these values within the caption of Table 5. The estimates of  $\beta_1$  and  $\phi$  for the GARCH(1,1) and GJR-GARCH(1,1) models are not much different under the Student-t and skewed Student-t choices for the innovation distribution for all three indices. The estimates for  $\phi$  under the two Bayesian nonparametric settings for  $F_e$ , DPM and IUM, for each GARCH-type model are marginally lower to the those under Student-t and skewed Student-t innovations. In terms of the IUM model the estimates of  $\beta_1$  are actually similar whereas that of DPM are lower. There is some disparity across both parametric and Bayesian nonparametric models when it comes to the leverage coefficient  $\beta_2$ . If we set the parametric models as the benchmarks, then IUM seems to slightly under estimate for the FTSE 100 and Eurostoxx 50 and over-estimate for the S&P 500, whereas the DPM largely under estimates  $\beta_2$  for the FTSE 100 and Eurostoxx 50, but provides similar a estimate for the S&P 500. The EGARCH(1,1) estimates of  $\beta_1$ , and  $\beta_2$

vary across the distributional choices for the innovations and across indices, whereas the estimates of  $\phi$  are more similar. Looking at the estimates of  $\phi$  under the two SV(1)-t models and the SV(1)-DPM, we would say that those are close to the corresponding estimates of the EGARCH(1,1) models.

We also provide in-sample estimates of the kurtosis  $\varkappa_{y_t}$  of the unconditional distribution of  $y_t$ . We were able to check for the existence of a 4<sup>th</sup> moment and then proceed to calculate  $\varkappa_{y_t}$  for all GARCH-type set ups with the EGARCH(1,1) being the exception. In EGARCH models the exponential growth of the conditional variance of  $y_t$  changes with the level of shocks, which leads to the explosion of the unconditional variance of  $y_t$  when the probability of extreme shocks is large. This means that the existence of the unconditional variance of  $y_t$  (in the EGARCH case) depends on the choice of innovations distribution. We have chosen the Student-t and skewed Student-t to model the innovations in the parametric models and under these two choices the unconditional variance of  $y_t$  does not exist. The formulas used to check for the existence of the fourth moment, and calculate  $\varkappa_{y_t}$  for the remaining GARCH-type and SV set ups are included in Appedix 3. These formulas were extracted from Jondeau and Rockinger (2003), Carnero-Angeles et al. (2004), and Verhoeven and McAleer (2004). All these formulas are based on the estimates of the volatility coefficients, and the estimated kurtosis of the innovation distribution,  $\varkappa_\epsilon$ . We used the estimates of the degrees of freedom and the skewness parameter (where applicable) to calculate  $\varkappa_\epsilon$ . For the Bayesian nonparametric GARCH-type models, the IUM and the DPM, the estimated kurtosis of the innovation distribution  $\varkappa_\epsilon$  is that of the point estimate  $\widehat{F_\epsilon(\cdot)}$  of  $F_\epsilon$ . It is not possible to calculate an estimate of the unconditional kurtosis of  $y_t$  under the  $\pi$ DDP model

of Griffin and Steel (2006) and under the SV(1)-DPM of Jensen and Maheu (2010). For the former model this is because the volatility is modelled using an ordered dependent DP, and thus no volatility coefficient estimates exist, and for the latter model the innovation distribution is not specified. The estimates of the unconditional kurtosis of  $y_t$  are displayed in Table 6 and the estimates of the kurtosis of the innovation distribution are displayed in Table 7. There are cases where the relevant condition for the existence of a fourth moment was not met and this occurs mostly with the Eurostoxx 50. For this reason the unconditional kurtosis of  $y_t$  is considered unbounded and cannot be estimated. We were however able to calculate  $\hat{\kappa}_{y_t}$  for most models for the other two indices, the S&P 500 and FTSE 100. All GARCH-type models with DPM innovations tend to over-estimate  $\hat{\kappa}_{y_t}$  for all indices. For the FTSE 100 we have three estimates that are close to the empirical value, those of the GARCH(1,1)-IUM, GJR-GARCH(1,1)-IUM and SV(1)-t (with and without leverage). The estimates of the GARCH(1,1)-IUM model are closest to the empirical values of kurtosis displayed in Table 3 for the S&P 500 and Eurostoxx 50. These results support our motivation for using the IUM to model the innovation distribution.

Finally, in Table 8 we provide in-sample estimates of the skewness parameter  $\lambda$  for all GARCH-type models with skewed Student-t, DPM and IUM innovations. Overall, the estimates of  $\lambda$  for models with IUM innovations are closest to the sample estimates for all indices.

## 6 Conclusions

This paper introduces a new Bayesian semiparametric model for the conditional distribution of daily stock index returns. The innovation distribution is modelled nonparametrically with an infinite mixture of scaled uniform distributions (IUM) rather than an infinite mixture of normals, as is the case with the DPM. We replaced the normal kernel with the scaled uniform kernel in order to capture the uni-modality, asymmetry and heavy-tailed behaviour of the conditional return distribution. We used a stick-breaking prior (SBP), instead of a Dirichlet process prior (DP) to model the unknown number of mixture clusters, because it has the flexibility to generate more clusters with non-negligible weights that could account for the heavy tails of the conditional return distribution. The uniform kernel together with the stick-breaking prior can absorb extreme returns via the heavy tails and avoid the risk of creating a mode at those extreme points, as may be the case with the DPM model. We developed an efficient MCMC based on the slice-efficient sampler introduced in Kalli et al. (2011) which samples the volatility coefficients as a block using adaptive MH. In our simulated study based on a GARCH(1,1) model our IUM has the best fit and predictive performance when compared to the DPM.

We accounted for the ‘leverage effect’ using both a GJR-GARCH(1,1) and an EGARCH(1,1) and tested our IUM on three samples of daily index returns taken from the S&P 500, FTSE 100 and Eurostoxx 50. We compared our model to GARCH(1,1), GJR-GARCH(1,1), and EGARCH(1,1) with Student-t and skewed Student-t, and DPM innovations, and found that its’ predictive performance in terms of extreme returns is better. It also accounts for kurtosis and skewness

more accurately. We came to the same conclusion when we compared the IUM to an SV(1) with Student-t innovations (with and without leverage), the SV(1)-DMP of Jensen and Maheu (2010) and  $\pi$ -DDP Griffin and Steel (2006). The over-arching conclusion is that the IUM model developed in this paper is competitive and can lead to improved inferences and predictions in the context of returns data analysis.

## References

- Atchade, Y. and Rosenthal, J. (2005). On adaptive Markov chain Monte Carlo algorithms., *Bernoulli* 11: 815–828.
- Bai, X., Russell, J. and Tiao, G. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations, *Journal of Econometrics* 114(2): 349–360.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31: 307–327.
- Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics* 69: 542–547.
- Bollerslev, T. (2008). Glossary to ARCH (GARCH), *Technical Report 2008-49*, Center of Research in Econometric Analysis of Time Series, School of Economics and Management, University of Aarhus.
- Carnero-Angeles, M., Pena, D. and Ruiz, E. (2004). Persistence and Kurtosis in

- GARCH and Stochastic Volatility Models, *Journal of Financial Econometrics* **2**(2): 319–342.
- Chen, C. and So, M. (2006). On threshold heteroscedastic model, *International Journal of Forecasting* **22**: 73–89.
- Chen, Q., Gerlach, R. and Lu, Z. (2011). Bayesian Value at Risk and expected shortfall forecasting via the asymmetric Laplace distribution, *Computational Statistics and Data Analysis* p. Forthcoming.
- Chib, S. and Hamilton, B. (2002). Semiparametric Bayes analysis of longitudinal data treatment models., *Journal of Econometrics* **110**: 67–89.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues., *Quantitative Finance* **1**(2): 223–236.
- Delatola, E. and Griffin, J. (2011). Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility, *Bayesian Analysis* **6**: 901–926.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K Inflation, *Econometrica* **987–1008**.
- Feller, W. (1957). *Introduction to Probability Theory and its Applications*, Vol. 1, John Wiley and Sons.
- Ferguson, T. (1973). A Bayesian Analysis of Some Nonparametric Problems, *The Annals of Statistics* **1**: 209–230.
- Fernandez, C. and Steel, M. F. J. (1998). On Bayesian Modelling of Fat Tails and Skewness, *Journal of the American Statistical Association* **93**: 359–371.

- Freedman, D. (1963). On the Asymptotic Behaviour of Bayes Estimates in the Discrete Case, *Annals of Mathematical Statistics* **34**: 1386–1403.
- Gallant, A. and Tauchen, G. (1989). Semi-nonparametric estimation of conditionally constrained heterogeneous processes: asset prices application, *Econometrica* **57**: 1091–1120.
- Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics* **41**: 337–348.
- Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *The Journal of Finance* **48**(5): 1779–1801.
- Griffin, J. E. and Steel, M. F. J. (2006). Order-Based Dependent Dirichlet Processes, *Journal of the American Statistical Association* **101**: 179–194.
- Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm., *Bernoulli* **7**: 223–242.
- Halmos, P. (1944). Random alms, *The Annals of Mathematical Statistics* **15**: 182–189.
- Hansen, B. E. (1994). Autoregressive conditional density estimation, *International Economic Review* **35**: 705–730.
- Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models., *Econometrica* **70**: 781–799.

- Hjort, N. L., Holmes, C., Müller, P. and Walker, S. (eds) (2010). *Bayesian Nonparametrics.*, Statistic and Probabilistic Mathematics, 1st edn, Cambridge University Press.
- Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for Stick-Breaking Priors, *Journal of the American Statistical Association* **96**: 161–173.
- Jacquier, E., Polson, N. and Rossi, P. E. (1994). Bayesian Analysis of Stochastic Volatility Models, *Journal of Business and Economic Statistics* **12**: 371–417.
- Jacquier, E., Polson, N. and Rossi, P. E. (2004). Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors, *Journal of Econometrics* **122**: 185–212.
- Jensen, M. (2004). Semiparametric Bayesian Inference of Long Memory Stochastic Volatility Models, *Journal of Time Series Analysis* **25**(6): 895–922.
- Jensen, M. J. and Maheu, J. M. (2010). Bayesian Semiparametric Stochastic Volatility Modeling, *Journal of Econometrics* **157**(2): 306–316.
- Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness and kurtosis: existence, persistence, and comovements., *Journal of Economic Dynamics and Control* **27**: 1699–1737.
- Kalli, M., Griffin, J. and Walker, S. (2011). Slice Sampling Mixture Models, *Statistics and Computing* **21**: 93–105.
- Kingman, J. (1974). Random Discrete Distributions, *Journal of the Royal Statistical Society, Series B.* **37**: 1–22.



- Leslie, D., Kohn, R. and Nott, D. (2007). A general approach to heteroscedastic linear regression, *Statistics and Computing* **17**: 131–146.
- Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates, *The Annals of Statistics* **12**: 351–357.
- Nelson, D. B. (1990). Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* **6**: 318–334.
- Poon, S. and Granger, C. (2003). Forecasting Volatility in Financial Markets: a review, *Journal of Economic Literature* **41**: 478–539.
- Roberts, G., Gelman, A. and Gilks, W. (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms, *Annals of Applied Probability* **7**: 110–120.
- Roberts, G. and Rosenthal, J. (2009). Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics* **18**: 349–367.
- Shahbaba, B. (2009). Discovering Hidden Structures using Mixture Models: Application to Nonlinear Time Series Processes, *Studies in Nonlinear Dynamics and Econometrics* **13**(5): <http://www.bepress.com/snnde/vol13/iss2/art5>.
- Taylor, S. (1982). *Time Series Analysis: Theory and Practice*, Vol. 1, North Holland, Amsterdam, chapter :Financial Returns Modelled by the Product of two Stochastic Processes- a study of the daily sugar prices 1961-1975.
- Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction.*, Princeton University Press.

Theodossiou, P. (1998). Financial Data and the Skewed Generalized T Distribution, *Management Science* 44(12): 1650–1661.

Tsay, R. S. (2005). *Analysis of Financial Time Series*, Probability and Statistics, 2nd edn, Wiley.

Verhoeven, P. and McAleer, M. (2004). Fat tails and asymmetry in financial volatility models, *Mathematics and Computers in Simulation* 64: 351–361.

## Appendix 1

In Section 2 we introduced a method of specifying the parameters  $(a_j, b_j)$  of the SBP for the mixture weights  $w_j$ . In this Appendix we discuss this method in more detail.

If  $w_j \sim SBP(a_j, b_j)$  then

$$\xi_j = \mathbb{E}[w_j] = \frac{a_j}{a_j + b_j} \prod_{l < j} \left(1 - \frac{a_l}{a_l + b_l}\right) \text{ for } j > l.$$

Let  $\tau_j = a_j/(a_j + b_j)$ , we then require the sequence  $(\tau_j)$  to satisfy the following,

$$\tau_1 = \xi_1, \text{ and for } j > 1, \tau_j \prod_{l < j} (1 - \tau_l) = \xi_j.$$

For  $j > 1$ , set  $\tau_j = \left(1 - \sum_{l < j} \xi_l\right)^{-1} \xi_j$ . We know that the  $\sum_l \xi_l = 1$  therefore  $\xi_j < 1 - \sum_{l < j} \xi_l$  and  $\tau_j < 1$ . At this stage we need to check that  $\sum_{j=1}^{\infty} \log(1 + a_j/b_j) = +\infty$ , a condition to ensure that  $F$  is a random measure, see Ishwaran and James (2001). As  $a_j/b_j = \tau_j/(1 - \tau_j)$  and  $\log(1 + \tau_i/(1 - \tau_j)) = -\log(1 - \tau_j)$ , we have,

$$\sum_{j=1}^{N_\tau} -\log(1 - \tau_j) = -\log \prod_{j=1}^{N_\tau} (1 - \tau_j) = -\log \left(1 - \sum_{j=1}^{N_\tau} \xi_j\right) \rightarrow +\infty,$$

therefore  $F$  is a random measure.

We center this process over a distribution for the weights by choosing  $E[w_j] = \xi_j$ , where  $\xi_j$  is  $Pr(X = j)$  for a random variable  $X$  with a discrete distribution on  $1, 2, 3, \dots$ . The random variable  $X$  is given a Beta-Geometric distribution i.e.  $Pr(X = j) = p(1 - p)^{j-1}$  where  $p \sim \text{Be}(\alpha_1, \alpha_2)$ . This yields

$$\xi_j = \frac{\Gamma(\alpha_1 + \alpha_2)\Gamma(\alpha_1 + 1)\Gamma(\alpha_2 + j - 1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_1 + \alpha_2 + j)},$$

which allows us to control the number of non-negligible weights by choosing the values of  $(\alpha_1, \alpha_2)$ .

To understand how the variance of  $w_j$  affects the weight decay we re-parameterise  $a_j$  and  $b_j$  as  $a_j = r_j\tau_j$  and  $b_j = r_j(1 - \tau_j)$ . In order to specify  $r_j$  we consider

$$E(w_j^2) = \xi_j \frac{1 + r_j\tau_j}{1 + r_j} \prod_{l < j} \left(1 - \frac{r_l\tau_l}{1 + r_l}\right),$$

and write  $r_l/(1 + r_l) = q_l$ . This provides the expression for the variance of  $w_j$ ,

$$\text{Var}(w_j) = \xi_j [(1 - q_j) + q_j\tau_j] \prod_{l < j} (1 - q_l\tau_l) - \xi_j^2.$$

The choice of  $q_j$  is what controls the variance of  $w_j$  and determines how close in

probability the weights are to  $\xi_j$ . We choose  $q_j$  to satisfy  $q_j = (1 - \tau_j)^{-1} \left[1 - \frac{\text{Var}(w_j) + \xi_j^2}{\xi_j \prod_{l < j} (1 - q_l\tau_l)}\right]$ .

Since  $0 < q_j < 1$ , we need

$$0 < 1 - \frac{\text{Var}(w_j) + \xi_j^2}{\xi_j \prod_{l < j} (1 - q_l\tau_l)} < 1 - \tau_j,$$

and hence

$$\tau_j \xi_j \prod_{l < j} (1 - q_l\tau_l) - \xi_j^2 < \text{Var}(w_j) < \xi_j \prod_{l < j} (1 - q_l\tau_l) - \xi_j^2.$$

One particular idea, which is used in the numerical illustrations, is to take large variances, amounting to a non-informative prior. This implies  $q_j$  is chosen to be small, but not zero, and hence one could set  $\xi_j \prod_{l < j} (1 - q_l\tau_l) - \xi_j^2 = c_\xi$ ,

for some small  $c_\xi$ , for all  $j$ . This follows since  $\text{Var}(w_j) < \xi_j(1 - \xi_j)$  and we obtain this limit as  $q_j \downarrow 0$ .

## Appendix 2

In both the SBP with normals and DPM case the hierarchical set up is:

$$\epsilon_t \sim N(\mu_{dt}, \sigma_{dt}^2) \text{ for } t = 1, \dots, n$$

$$\Pr(d_t = j) = w_j \text{ for } j = 1, 2, \dots$$

$$\mu_j, \sigma_j^2 \sim G_0(\cdot) \text{ for } j = 1, 2, \dots$$

$$\text{where } G_0(\cdot) \text{ is } \mu_j | \sigma_j^2 \sim N(m, \frac{\sigma_j^2}{\gamma}) \text{ and } \frac{1}{\sigma_j^2} \sim \text{Ga}(a_\sigma, b_\sigma)$$

$$w_1 = v_1 \text{ and } w_j = (1 - v_1) \dots (1 - v_{j-1}) v_j = v_j \prod_{l < j} (1 - v_l)$$

$$v_j \sim \text{Be}(a_j, b_j) \text{ for the SBM, and } v_j \sim \text{Be}(1, c) \text{ for the DPM.}$$

For our simulation example we have set  $m = 0.0001$ ,  $\gamma = 0.1$ ,  $a_\sigma = 0.05$ , and  $b_\sigma = 0.05$ . The precision parameter for the DPM is set  $c = 1$  and the  $a_j, b_j$  of the SBP are generated from the geometric-beta model (described in Section 2) with parameters  $\alpha_1 = 1$ ,  $\alpha_2 = 6$ .

We use the MCMC algorithm described in Section 3 for all models. For both SBP with the normal set up and DPM the joint posterior distribution is proportional to

$$\prod_{t=1}^n \mathbf{1}(u_t < \zeta_{d_t}) w_{d_t} \zeta_{d_t}^{-1} N(\epsilon_t | \mu_{d_t}, \sigma_{d_t}^2).$$

Recall that  $\zeta$  and  $v$  are conditionally independent and so the sampling steps are:

- $\pi(v_j) \propto \text{beta}(v_j; a'_j, b'_j)$  for the SBM where  $a'_j$  and  $b'_j$  are defined just like in Section 3, and  $\pi(v_j) \propto \text{beta}(v_j; 1', c')$  for the DPM where  $1' = 1 + \sum_{t=1}^n \mathbf{1}(d_t = j)$  and  $c' = c + \sum_{t=1}^n \mathbf{1}(d_t > j)$ .

- $\pi(\mu_{d_t} | \dots) \sim \mathbf{N}\left(\frac{\sum_{t:d_t=j} \epsilon_t + m\gamma}{n_j + \gamma}, \frac{\sigma_{d_t}^2}{n_j + \gamma}\right)$
- $\pi\left(\frac{1}{\sigma_{d_t}^2} | \dots\right) \sim \mathbf{Ga}\left(a_\sigma + \frac{n_j}{2} + \frac{1}{2}, b_\sigma + \frac{\sum_{t:d_t=j} (\epsilon_t - \mu_{d_t})^2}{2} + \frac{\gamma(\mu_{d_t} - m)^2}{2}\right)$
- $\pi(u_t | \dots) \propto \mathbf{1}(0 < u_t < \zeta_{d_t})$ .
- $\Pr(d_t = \kappa | \dots) \propto \mathbf{1}(\kappa : \zeta_\kappa > u_t) w_\kappa \zeta_\kappa^{-1} \mathbf{N}(\epsilon_t | \mu_{d_t}, \sigma_{d_t}^2)$

The random walk MH described in Section 3 is used to sample the volatility coefficients. What changes in the case of the normal setup is the likelihood and as a consequence the simplified likelihood used the MH sampler. The likelihood is proportional to

$$\prod_{t=1}^n \frac{e^{(\epsilon_t - \mu_{d_t})^2 / 2\sigma_{d_t}^2}}{\sqrt{2\pi\sigma_{d_t}^2}} \text{ where } \epsilon_t = \frac{y_t}{h_t}$$

The simplified likelihood is obtained by integrating out the  $\mu_{d_t}$  over  $(-\infty, \infty)$  and then the  $\sigma_{d_t}^2$  over  $(0, \infty)$ , and it is proportional to

$$\frac{1}{\prod_{t=1}^n h_t} \prod_{j=1}^{N_c} \Upsilon_j$$

where  $N_c$  is the number of full clusters,  $n_j$  the size of cluster  $j$  and

$$\Upsilon_j = \frac{\gamma^{\frac{1}{2}} b_\sigma^{a_\sigma} \Gamma(a_\sigma + \frac{n_j}{2})}{(2\pi)^{\frac{n_j}{2}} (\gamma + n_j)^{\frac{1}{2}} \Gamma(a_\sigma) \left( b_\sigma + \frac{1}{2} (\gamma m^2 + \sum_{j=1}^{n_j} \epsilon_j^2 - \frac{(\gamma m + \sum_{j=1}^{n_j} \epsilon_j^2)^2}{\gamma + n_j}) \right)^{a_\sigma + \frac{n_j}{2}}}$$

### Appendix 3

The GARCH(1,1) condition for the existence of a fourth moment is:

$$(\beta_1 + \phi)^2 + \beta_1^2 (\varkappa_\epsilon - 1) < 1.$$

If this condition is satisfied then the unconditional kurtosis of  $y_t$  is

$$\varkappa_{y_t} = \varkappa_\epsilon \left[ 1 - \frac{\beta_1^2 (\varkappa_\epsilon - 1)}{1 - (\beta_1 + \phi)^2} \right]^{-1},$$

where  $\beta_1$  and  $\phi$  are the coefficients of  $y_{t-1}^2$  and  $h_{t-1}^2$  respectively, and  $\varkappa_\epsilon$  is the kurtosis of the innovation distribution (see Carnero-Angeles et al. (2004)).

The GJR-GARCH(1,1) condition for the existence of a fourth moment is

$$\phi^2 + 2\beta_1\phi + \phi\beta_2 + \varkappa_\epsilon\beta_1\beta_2 + \varkappa_\epsilon\beta_1^2 + \frac{1}{2}\beta_2^2\varkappa_\epsilon < 1$$

If this condition is satisfied then the unconditional kurtosis of  $y_t$  is

$$\varkappa_{y_t} = \varkappa_\epsilon \frac{1 - (\phi^2 + 2\beta_1\phi + \phi\beta_2 + \beta_1\beta_2 + \beta_1^2 + \frac{1}{4}\beta_2^2)}{1 - (\phi^2 + 2\beta_1\phi + \phi\beta_2 + \varkappa_\epsilon\beta_1\beta_2 + \varkappa_\epsilon\beta_1^2 + \frac{1}{2}\beta_2^2\varkappa_\epsilon)},$$

where  $\beta_1, \beta_2$  and  $\phi$  are the coefficients of  $y_{t-1}^2, y_{t-1}^2 I_{t-1}$  and  $h_{t-1}^2$  respectively, and  $\varkappa_\epsilon$  is the kurtosis of the innovation distribution (see Verhoeven and McAleer (2004)).

In our illustrations we are using an SV(1) set up for the volatility, that is  $\log h_t^2 = \beta_0 + \phi \log h_{t-1}^2 + \eta_t$ , where  $\eta \sim N(0, \sigma_\eta^2)$ . In this case if the kurtosis of the innovations' distributions,  $\varkappa_\epsilon$ , is finite, the condition for the existence of the unconditional kurtosis of  $y_t$ ,  $\varkappa_{y_t}$ , is the stationarity condition, that is  $|\phi| < 1$ .

Then

$$\varkappa_{y_t} = \varkappa_\epsilon \exp(\sigma_h^2) \text{ with } \sigma_h^2 = \frac{\sigma_\eta^2}{(1 - \phi^2)}.$$

(see Carnero-Angeles et al. (2004))

The formula for calculating the kurtosis of the student-t distribution is

$$\varkappa_\epsilon = 3 + \frac{6}{df - 4} \text{ for } df > 4,$$

where  $df$  are the degrees of freedom. Calculating the kurtosis of the skewed student-t distribution is more challenging. As with any distribution the kurtosis of the innovations' distribution is given by

$$\varkappa_\epsilon = \frac{\mathbf{E}(\epsilon^4) - 4\mathbf{E}(\epsilon)E(\epsilon^3) + 6[\mathbf{E}(\epsilon)]^2\mathbf{E}(\epsilon^2) - 3[\mathbf{E}(\epsilon)]^4}{[\mathbf{Var}(\epsilon)]^2}.$$

We use the formulas in Jondeau and Rockinger (2003) to calculate these moments. The formulas are,

$$E(\epsilon) = 4\lambda c_\epsilon \frac{df - 3}{df - 1}$$

$$E(\epsilon^2) = 1 + 3\lambda^2$$

$$E(\epsilon^3) = 16c_\epsilon \lambda (1 + \lambda^2) \frac{(df - 2)^2}{(df - 1)(df - 3)}$$

$$E(\epsilon^4) = 3 \frac{(df - 2)}{(df - 4)} (1 + 10\lambda^2 + 5\lambda^4),$$

where  $c_\epsilon = \frac{\Gamma(\frac{df+1}{2})}{\sqrt{\pi(df-2)\Gamma(\frac{df}{2})}}$ , and  $\lambda$  is the skewness parameter.

	S&P 500			FTSE 100			Eurostoxx 50		
	LPS	LPTS-01	LPTS-05	LPS	LPTS-01	LPTS-05	LPS	LPTS-01	LPTS-05
Garch(1,1)-t	1.324	7.108	4.069	1.360	5.706	3.572	1.696	5.729	4.303
Garch(1,1)-skewed t	1.314	7.653	4.409	1.351	5.799	3.650	1.697	6.403	4.377
GJR-Garch(1,1)-t	1.322	7.111	4.099	1.349	5.242	3.386	1.658	5.643	3.981
GJR-Garch(1,1)-skewed t	1.314	8.039	4.588	1.337	5.402	3.490	1.668	6.329	4.359
EGarch(1,1)-t	1.328	6.076	3.719	1.334	5.459	3.486	1.668	5.098	3.285
EGarch(1,1)-skewed t	1.310	8.040	4.427	1.328	5.555	3.554	1.666	6.241	4.331
Garch(1,1)-DPM	1.313	5.961	4.039	1.330	5.712	3.588	1.667	5.109	3.155
GJR-Garch(1,1)-DPM	1.311	6.017	4.059	1.327	5.831	3.600	1.768	5.211	4.112
Egarch(1,1)-DPM	1.926	6.735	4.586	1.355	5.353	3.412	1.693	6.012	4.312
Garch(1,1)-IUM	1.302	5.943	3.612	1.326	5.679	3.501	<b>1.655</b>	<b>4.873</b>	<b>2.989</b>
GJR-Garch(1,1)-IUM	1.319	<b>5.650</b>	<b>3.598</b>	<b>1.319</b>	<b>4.897</b>	<b>3.094</b>	1.695	5.132	3.877
Egarch(1,1)-IUM	1.323	6.129	3.815	1.332	5.301	3.379	1.683	5.548	4.121
SV(1)-t	1.330	8.712	4.835	1.349	5.627	4.126	1.754	4.891	3.321
SV(1)-t (leverage)	1.325	9.023	4.879	1.343	5.705	3.567	1.748	5.113	3.435
SV-DPM	1.313	7.231	4.811	1.349	5.963	3.712	1.701	6.667	4.867
$\pi$ -DDP	<b>1.301</b>	5.941	3.934	1.409	5.810	3.231	1.687	5.008	3.8143

Table 4: Log predictive scores and log predictive tail scores at 1% and 5% for S&amp;P 500, FTSE 100, and Eurostoxx 50.



	S&P 500				FTSE 100				Eurostoxx 50			
	$\beta_0$	$\beta_1$	$\beta_2$	$\phi$	$\beta_0$	$\beta_1$	$\beta_2$	$\phi$	$\beta_0$	$\beta_1$	$\beta_2$	$\phi$
Garch(1,1)-t	0.022 (0.006,0.038)	0.042 (0.022,0.062)	-	0.934 (0.903,0.965)	0.011 (0.003,0.020)	0.100 (0.078,0.122)	-	0.897 (0.875,0.920)	0.012 (0.002,0.022)	0.084 (0.061,0.108)	-	0.912 (0.889,0.936)
Garch(1,1) skew t	0.023 (0.022,0.023)	0.043 (0.0428,0.0432)	-	0.932 (0.931,0.933)	0.011 (***)	0.095 (0.0948,0.0952)	-	0.902 (0.9018,0.9022)	0.012 (0.0116,0.0124)	0.080 (0.0796,0.0804)	-	0.915 (0.913,0.917)
GJR-Garch(1,1)-t	0.030 (0.01,0.05)	0.033 (0.009,0.057)	0.035 (0.002,0.068)	0.918 (0.883,0.953)	0.014 (0.008,0.020)	0.001 (***)	0.144 (0.111,0.177)	0.918 (0.898,0.938)	0.017 (0.009,0.025)	0.0001 (***)	0.156 (0.111,0.201)	0.914 (0.889,0.940)
GJR-Garch(1,1) skew t	0.031 (0.030,0.032)	0.034 (***)	0.035 (0.033,0.037)	0.916 (0.914,0.918)	0.013 (***)	0.0001 (***)	0.138 (0.137,0.139)	0.922 (0.9218,0.9222)	0.016 (***)	0.0001 (***)	0.155 (0.154,0.157)	0.914 (0.9136,0.9144)
Egarch(1,1)-t	-0.019 (-0.033,-0.005)	0.176 (0.115,0.237)	-0.108 (-0.145,-0.071)	0.910 (0.871,0.950)	0.0001 (***)	0.122 (0.089,0.155)	-0.114 (-0.138,-0.091)	0.985 (0.981,0.989)	0.005 (0.001,0.010)	0.108 (0.069,0.147)	-0.140 (-0.167,-0.113)	0.985 (0.981,0.989)
Egarch(1,1) skew t	0.004 (***)	0.107 (0.106,0.108)	-0.033 (-0.0334,-0.0326)	0.982 (***)	0.003 (***)	0.116 (***)	-0.112 (***)	0.985 (***)	0.007 (***)	0.104 (***)	-0.143 (-0.144,-0.142)	0.985 (***)
Garch(1,1)-DPM	0.013 (0.002,0.056)	0.021 (0.004,0.074)	-	0.922 (0.866,0.952)	0.007 (0.001,0.016)	0.061 (0.016,0.106)	-	0.899 (0.866,0.919)	0.000 (***)	0.000 (***)	-	0.837 (0.797,0.918)
GJR-Garch(1,1)-DPM	0.032 (-0.007,0.119)	0.030 (0.005,0.079)	0.034 (0.005,0.115)	0.900 (0.828,0.937)	0.100 (0.034,0.162)	0.005 (0,0.019)	0.012 (0.003,0.020)	0.916 (0.890,0.932)	0.005 (0.001,0.019)	0.000 (***)	0.001 (0.000,0.002)	0.909 (0.876,0.928)
Egarch(1,1)-DPM	-0.041 (-0.114,-0.013)	0.137 (0.061,0.254)	-0.040 (-0.062,-0.001)	0.978 (0.952,0.990)	-0.173 (-0.308,-0.109)	0.188 (0.100,0.259)	-0.003 (-0.012,-0.001)	0.985 (0.978,0.990)	0.000 (-0.167,0.000)	0.000 (0.000,0.082)	-0.382 (-0.600,-0.029)	0.978 (0.965,0.988)
Garch(1,1)-IUM	0.025 (0.001,0.176)	0.045 (0.008,0.095)	-	0.918 (0.851,0.958)	0.003 (0,0.009)	0.032 (0.003,0.100)	-	0.903 (0.890,0.915)	0.002 (0,0.003)	0.012 (0.009,0.019)	-	0.920 (0.895,0.932)
GJR-Garch(1,1)-IUM	0.060 (0.013,0.150)	0.044 (0.006,0.108)	0.072 (0.015,0.126)	0.890 (0.790,0.920)	0.014 (0.004,0.024)	0.004 (0,0.013)	0.111 (0.030,0.163)	0.912 (0.902,0.923)	0.001 (0,0.007)	0.0002 (0,0.002)	0.135 (0.009,0.165)	0.926 (0.907,0.941)
Egarch(1,1)-IUM	0.08 (0.001,0.121)	0.192 (0.126,0.283)	-0.078 (-0.123,-0.033)	0.960 (0.917,0.988)	0.0003 (-0.0032,0.0271)	0.132 (0.082,0.190)	-0.145 (-0.192,-0.099)	0.986 (0.984,0.989)	-0.057 (-0.102,0.004)	0.089 (0.029,0.199)	-0.132 (-0.208,-0.040)	0.983 (0.979,0.989)
SV(1)-t	-0.015 (-0.029,-0.004)	-	-	0.962 (0.934,0.980)	0.00 (-0.006,0.006)	-	-	0.984 (0.976,0.992)	0.002 (-0.006,0.011)	-	-	0.988 (0.979,0.995)
SV(1)-t leverage	-0.005 (-0.015,0.002)	-	-	0.984 (0.9663,0.996)	0.007 (0.003,0.012)	-	-8 (0.978,0.990)	0.985 (0.978,0.990)	0.011 (0.006,0.017)	-	-	0.986 (0.981,0.991)
SV(1)-DPM	-	-	-	0.984 (0.963,0.997)	-	-	-	0.989 (0.979,0.997)	-	-	-	0.992 (0.982,0.999)

Table 5: Estimates of the volatility coefficients for S&P 500, FTSE 100, and Eurostoxx 50. For the parametric models the estimates with their 95% confidence intervals are shown. For the Bayesian nonparametric models the posterior medians with the 95% credible intervals are shown. Note that \*\*\* signifies either a standard error virtually close to zero or bounds very close to zero and hence 95% confidence intervals (or 95% credible intervals) were not calculated. The correlation coefficient estimates for the SV(1)-t with leverage were  $-0.240$ ,  $-0.758$ , and  $-0.819$  for each index respectively.

	S&P 500	FTSE 100	Eurostoxx 50
Garch(1,1)-t	3.71	***	***
Garch(1,1) skew t	6.27	***	***
Garch(1,1)-DPM	55.37	49.47	20.01
Garch(1,1)-IUM	85.70	9.63	7.68
GJR-Garch(1,1)-t	6.33	***	***
GJR-Garch(1,1) skew t	6.19	***	***
GJR-Garch(1,1)-DPM	***	40.04	21.10
GJR-Garch(1,1)-IUM	***	10.69	***
SV(1)-t	5.60	8.34	10.94
SV(1)-t-leverage	5.82	7.64	8.65

Table 6: Estimates of unconditional kurtosis,  $\varkappa_{y_t}$ , of  $y_t$ . The \*\*\* signify failure to meet the conditions for the existence of a 4<sup>th</sup> moment, resulting in unbounded kurtosis (see Appendix 2)

	S&P 500	FTSE 100	Eurostoxx 50
Garch(1,1)-t	3.38	3.55	4.08
Garch(1,1) skew t	5.27	3.65	4.14
Garch(1,1)-DPM	46.00	15.68	20.01
Garch(1,1)-IUM	25.99	9.01	7.62
GJR-Garch(1,1)-t	5.10	3.50	3.60
GJR-Garch(1,1) skew t	5.01	3.51	3.65
GJR-Garch(1,1)-DPM	48	38.68	21.1
GJR-Garch(1,1)-IUM	13.56	5.03	7.89
SV(1)-t	3.67	3.19	3.24
SV(1)-t-leverage	4.00	3.20	3.24

Table 7: Estimates of the kurtosis of the innovation distribution,  $\varkappa_\epsilon$ . (see Appendix 2)

	S&P 500	FTSE 100	Eurostoxx 50
<i>Sample estimates</i>	<i>-4.13 (-0.022)*</i>	<i>-0.124</i>	<i>0.099</i>
Garch(1,1)-skew t	-0.016	-0.111	-0.091
GJR-Garch(1,1) skew t	-0.011	-0.127	-0.120
Egarch(1,1) skew t	-0.013	-0.128	-0.129
Garch(1,1)-DPM	-0.009	-0.093	0.000
GJR-Garch(1,1)-DPM	-0.010	-0.108	-0.005
Egarch(1,1)-DPM	-0.010	-0.110	-0.011
Garch(1,1)-IUM	-0.025	-0.125	0.029
GJR-Garch(1,1)-IUM	-0.019	-0.130	0.078
EGarch(1,1)-IUM	-0.028	-0.131	0.091

Table 8: Estimates of skewness parameter,  $\lambda$ . Comparisons are for GARCH-type models with skewed-t, DPM, and IUM innovations. Sample estimates are in italics. \*In parenthesis: skewness estimate when the extreme outlier is removed.