A SEMIPARAMETRIC BAYESIAN APPROACH TO THE ANALYSIS OF FINANCIAL TIME SERIES WITH APPLICATIONS TO VALUE AT RISK ESTIMATION

M. Concepción Ausin[1], Pedro Galeano[1] and Pulak Ghosh[2]

Abstract

Financial time series analysis deals with the understanding of data collected on financial markets. Several parametric distribution models have been entertained for describing, estimating and predicting the dynamics of financial time series. Alternatively, this article considers a Bayesian semiparametric approach. In particular, the usual parametric distributional assumptions of the GARCH-type models are relaxed by entertaining the class of location-scale mixtures of Gaussian distributions with a Dirichlet process prior on the mixing distribution, leading to a Dirichlet process mixture model. The proposed specification allows for a greater exibility in capturing both the skewness and kurtosis frequently observed in financial returns. The Bayesian model provides statistical inference with finite sample validity. Furthermore, it is also possible to obtain predictive distributions for the Value at Risk (VaR), which has become the most widely used measure of market risk for practitioners. Through a simulation study, we demonstrate the performance of the proposed semiparametric method and compare results with the ones from a normal distribution assumption. We also demonstrate the superiority of our proposed semiparametric method using real data from the Bombay Stock Exchange Index (BSE-30) and the Hang Seng Index (HSI).

Keywords: Bayesian estimation; Deviance information criterion; Dirichlet process mixture; Financial time series; Location-scale Gaussian mixture; Markov chain Monte Carlo.

[1] Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Madrid, Spain, e-mail: concepcion.ausin@uc3m.es, pedro.galeano@uc3m.es
[2] Indian Institute of Management Bangalore, Bannerghatta Road, Bangalore, India, e-mail: pulak.ghosh@iimb.ernet.in

A Semiparametric Bayesian Approach to the Analysis of Financial Time Series with Application to Value at Risk Estimation

M. Concepción Ausín, Pedro Galeano and Pulak Ghosh

Abstract

Financial time series analysis deals with the understanding of data collected on financial markets. Several parametric distribution models have been entertained for describing, estimating and predicting the dynamics of financial time series. Alternatively, this article considers a Bayesian semiparametric approach. In particular, the usual parametric distributional assumptions of the GARCH-type models are relaxed by entertaining the class of location-scale mixtures of Gaussian distributions with a Dirichlet process prior on the mixing distribution, leading to a Dirichlet process mixture model. The proposed specification allows for a greater flexibility in capturing both the skewness and kurtosis frequently observed in financial returns. The Bayesian model provide statistical inference with finite sample validity. Furthermore, it is also possible to obtain predictive distributions for the Value at Risk (VaR), which has become the most widely used measure of market risk for practitioners. Through a simulation study, we demonstrate the performance of the proposed semiparametric method and compare results with the ones from a normal distribution assumption. We also demonstrate the superiority of our proposed semiparametric method using real data from the Bombay Stock Exchange Index (BSE-30) and the Hang Seng Index (HSI).

Keywords: Bayesian estimation; Deviance information criterion; Dirichlet process mixture; Financial time series; Location-scale Gaussian mixture; Markov chain Monte Carlo.

1 Introduction

Financial time series analysis gives practical and theoretical understanding of data collected on financial markets, such as stock and commodity prices, exchange rates or bond yields. Investors and financial managers need to understand the behavior of asset prices to have good expectations about future prices and the risks to which they will be exposed. Although forecasting is an essential component of any interesting activity, it is usually very difficult to obtain accurate predictions. Since, statistically the prediction methods inherently depends on the underlying distributions assumed, it appears to be more appropriate to gain insights into the assumed probability distributions to obtain better predictions for future values. Correctly specifying the distribution is also important as it provides with a measure of investment risk.

Financial data usually consists of a time series of prices of a certain asset for a given period of time. However, most of the financial analysis consider asset returns, which measures the relative changes in prices, as they have more attractive statistical properties. It is worth mentioning that modeling returns challenging as it exhibits several interesting and complicated features, usually called "stylized facts". First, returns appear to vary around the mean levels, which are close to zero. Thus, they are mean stationary, at least at certain periods of time. Second, the conditional volatility, which is usually measured through the conditional variance, is not constant over time. This is because, during periods of time in which economic crises, wars or political disorders happen, returns fluctuate strongly, while in tranquil periods, returns fluctuate weakly. This effect is called "volatility clustering". Third, returns usually take several large positive and negative values. More precisely, extreme negative returns are more frequent than extreme positive returns. As a consequence, the unconditional distribution of returns is known to be negatively skewed and heavy-tailed and thus a precise form of the tail is difficult to determine. Finally, although the serial correlation of returns is very small, if any, squared and absolute returns show strong serial correlation.

Engle (1982) introduced the autoregressive conditional heteroskedastic (ARCH) model to describe these stylized facts of financial returns. Since then, many alternative specifications have been proposed including the stochastic volatility (SV) model, proposed by Taylor (1982), the generalized autoregressive conditional heteroskedastic (GARCH) model proposed by Bollerslev (1986), the exponential generalized autoregressive conditional heteroskedastic (EGARCH) model proposed by Nelson (1991) and the GJR model proposed by Glosten, Jaganathan and Runkle (1993), among others. However, the GARCH model is by far the most popular model for estimating the dynamics of financial returns, mainly because of its ease in estimation and the availability of useful volatility forecasts. In order to get the probability distributions of future

returns inherited by GARCH-type models, it is necessary to specify the distribution of the innovations. The simpler and most routinely used assumption is that these innovations are normally distributed. While this assumption makes the model easy to implement, the accuracy of this assumption is questionable as it is well known that GARCH-type models with Gaussian innovations are inconsistent with many of the features of returns described before. Alternative popular approaches are the Student-t distribution (Bollerslev, 1987), the generalized error distribution, also called exponential power distribution (Nelson, 1991), and a mixture of two zero mean Gaussian distributions (Bai, Russell and Tiao, 2003). However, all these distributions are symmetric, and thus unable to capture the frequently observed skewness in the unconditional distribution of returns. Although, Gallant and Tauchen (1989) and Lee and Tse (1991) use Gram-Charlier expansions, Hansen (1994) and Fernández and Steel (1998) proposed the skewed Student-t distribution and Premaratne and Bera (2000) analyzed the Pearson IV distribution, skewed distributions in a GARCH model has not been explored much. Inference on GARCH-type models has been traditionally carried out in a classical setup by using the Gaussian quasi maximum likelihood estimator (QMLE). Despite the fact that this method has become popular, this approach presents some limitations, including slow numerical convergence; the fact that the sequence of parameter vectors converges to a maximum likelihood estimator only if a judicious choice of the starting value is made; and the underlying assumption of normality for the data set. These limitations, coupled with the non-existence of a measure of the standard errors for the estimates, serve to limit the methods applicability severely. Instead, we consider a Bayesian approach for inference. Bayesian approach has several advantages. The Bayesian model provide statistical inference with finite sample validity and any constraints on the model parameters can be easily incorporated through appropriate prior specifications. Moreover, using Markov Chain Monte Carlo (MCMC) one can get the joint posterior distributions of the model parameters (Ardia, 2008). Despite its attractiveness, the Bayesian analysis of GARCH model is relatively less explored. See for instance, the papers by Bauwens and Lubrano (1998), Müller and Pole (1998), Nakatsuma (2000), Vrontos, Dellaportas and Politis (2000), Kaufmann and Frühwirth-Schnatter (2002) and Ausin and Galeano (2007), among others.

In this article, we consider a semiparametric Bayesian approach to GARCH-type models that extends previous work in several ways. First, the usual parametric distributional assumptions on the innovations of GARCH-type models are relaxed by using a semiparametric Bayesian approach. In particular, instead of assuming a Gaussian, Student-t or a Gaussian mixture distribution, we broaden the class to a location-scale mixture of normal. The class of normal location-scale mixtures is quite broad and includes, the Gaussian, Student-t, logistic, double exponential, Cauchy and generalized hyperbolic distributions, among others. A

Dirichlet process (DP) prior (see, Fergusson, 1973) is assumed on the mixing distribution, resulting in a DP mixture (DPM) model, see, for instance, Antoniak (1974), Escobar (1994) and Escobar and West (1995). This specification, which shall be referred as GARCH-DPM model, allows for a greater flexibility in capturing the skewness and kurtosis of financial returns. Second, inference on the GARCH-DPM model has been developed using an MCMC algorithms which combines the ideas of retrospective sampling proposed in Papaspiliopoulos and Roberts (2008), the slice sampling of Walker (2007) and the Griddy-Gibbs sampling of Ritter and Tanner (1992). Data augmentation techniques and numerical integration methods are used such that all the conditional distributions are easy to sample without the need of finite truncation approaches nor accept/reject methods. Third, model selection among the parametric and semiparametric models is considered using the deviance information criterion (DIC) for mixture models recently proposed by Celeux, Forbes, Robert and Titterington (2006).

Extreme price movements in financial markets are unusual, but important. Recently, the large daily price movements have pointed out the need of reliable investment risk measures. Value at Risk (VaR) has become the most widely used measure of market risk. VaR indicates the potential loss associated with an unfavorable movement in market prices over a given time period at a certain confidence level. Statistically speaking, the VaR is a quantile of the conditional distribution of the returns. Thus, its calculation strongly depends on the assumption made for the innovation distribution. The proposed methodology, apart from offering a convenient specification of the innovation distribution, offers a natural way to introduce parameter uncertainty in the estimation and prediction of volatilities and VaR, usually ignored in the QMLE approach. Thus, the fourth contribution of this paper is to show how to obtain predictive distributions of in-sample volatilities and VaR, which provides a measure of precision for VaR estimates via predictive intervals. Also, predictive distributions of the future returns and volatilities can be obtained which are more informative than simple point forecasts.

The rest of this paper is organized as follows. Section 2 presents the motivating data set based on two Asian stock market indexes: the Bombay Stock Exchange Index (BSE-30) and the Hang Seng Index (HSI). Section 3 introduces the GARCH-DPM model and illustrate its flexibility in capturing the patterns exhibited by financial returns. Section 4 describes the MCMC algorithm to sample from exact posterior distribution of quantities of interest. Section 5 explains how to compute the DIC for the proposed semiparametric model. Section 6 addresses the problems of in-sample volatility estimation, out-of-sample return and volatility prediction and VaR estimation. Section 7 presents a brief Monte Carlo experiment which illustrates the accuracy in parameter estimation, prediction of volatilities and VaR estimation. Section 8 analyzes the

BSE-30 and HSI indexes using the proposed methodology. Finally, Section 9 concludes.

2 Motivating data: Asian stock market index

With the expansion of international financial links and the continued liberalization of cross-border cash flows, study of international stock markets have become increasingly prevalent. In particular, emerging markets such as India and China tend to have high volatile performance and experience economic crises more frequently than developed economies. As an illustration of the usefulness of the proposed semiparametric approach, we will analyze in this paper two important stock market indexes in Asia: the Bombay Stock Exchange Index and the Hang Seng Index. However, the applications of the proposed method are clearly beyond these examples.

The BSE-30 index is a value-weighted index composed of the 30 largest stocks, representative of various sectors, of the Bombay Stock Exchange. The HSI index is a freefloat-adjusted market capitalization-weighted stock market index composed of 45 companies of the Hong Kong Stock Exchange. Figures 1 and 2 show the time plot of the daily closing prices adjusted for dividends and splits of both indexes, for the period from January 2000 until December 2009, leading to 2474 and 2489 index levels, respectively. Note how the BSE-30 started a sudden fall at January, 2008 from the peak level of 20873, which continued until a sudden increase at March, 2009 from the peak level of 8160 until the end of the series. On the other hand, the HSI started a sudden fall at October, 2007 from the peak level of 31638, which continued until a sudden increase at October, 2008 from the peak level of 11015 until the end of the series. Figures 1 and 2 also show the time plot of the returns of the daily closing prices of the two indexes. Observe that the returns appear to vary more in the last part of the series, which is the period corresponding to the recent financial crisis. Also, note a large peak in the BSE-30, which corresponds to May 18, 2009, where this index increased a 16% primarily due to the victory of the United Progressive Alliance (UPA) in the Indian general elections. We can also observe the large peaks in the HSI returns corresponding to October 27 and 28, 2008, where this index first decreased and after increased around a 13.5%, primarily due to the fear of a world recession. Finally, Table 1 shows some summary statistics. Observe that both time series are left-skewed and the kurtosis are rather larger than 3, indicating that the distributions have higher peaks and heavier tails than a normal distribution with the same variance.

Thus, it seems that a Bayesian semiparametric model can be an adequate approach to better address these issues. With this purpose in mind, the next section introduces a generalized autoregressive heteroscedastic
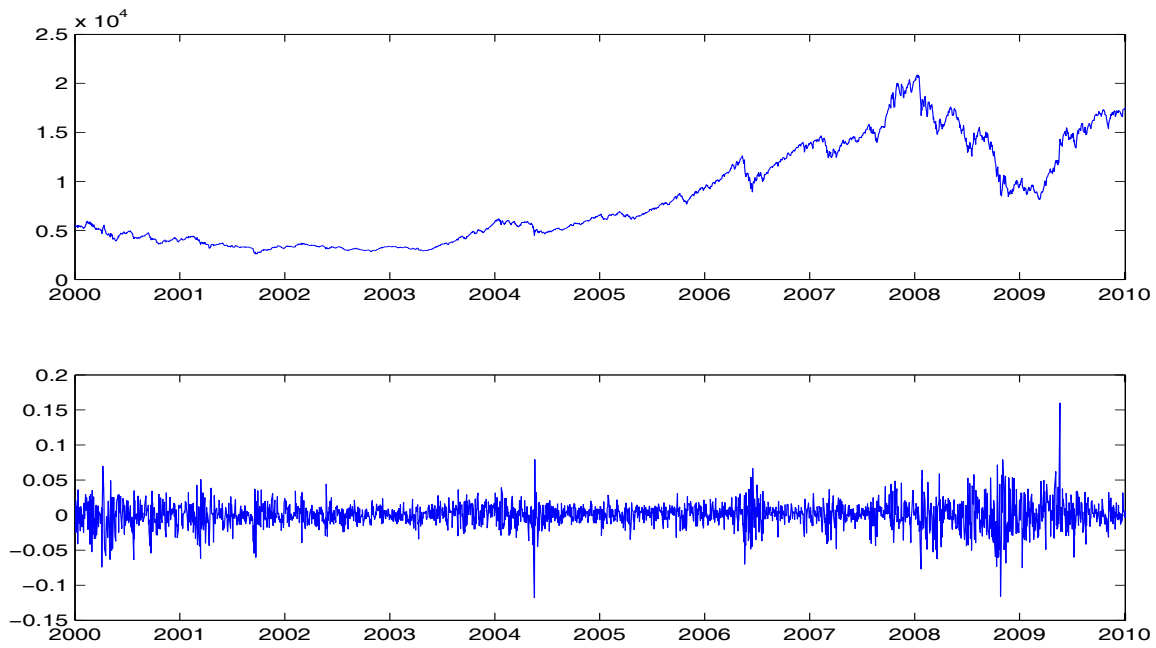
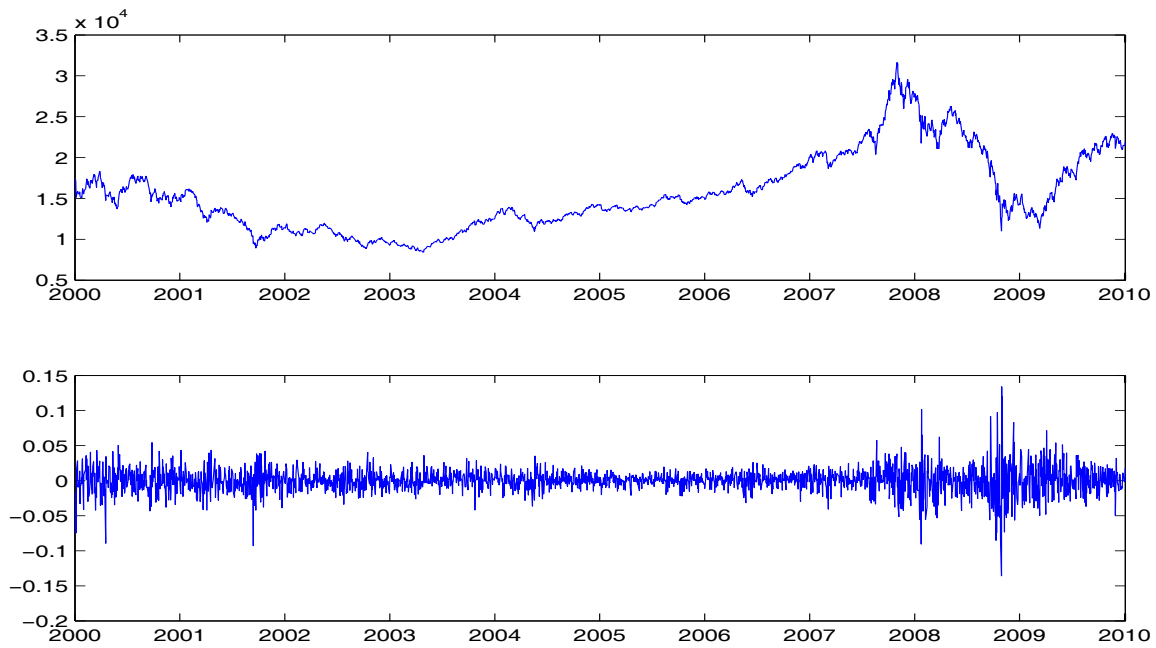Figure 1: Levels (top) and returns (bottom) of the Bombay Stock Exchange index.

Figure 2: Levels (top) and returns (bottom) of the Hang Seng index

Table 1: Summary statistics of the daily returns of the BSE-30 and HSI indexes.

	BSE-30	HSI
Mean	0.0477	0.0093
Standard deviation	1.7904	1.7077
Skewness	-0.1980	-0.0375
Kurtosis	8.7109	10.5942

model using Dirichlet Processes.

3 The Generalized Autoregressive Heterocedastic Dirichlet Process Mixture model

Among the class of volatility models, GARCH-type models are the most popular for describing financial returns because they are simple to estimate and provide with useful volatility forecasts. For these models, the volatility only depends on the past observations. For instance, in the standard GARCH(p, q) model, the return series is assumed to follow,

$$r_t = h_t^{1/2}\epsilon_t, \tag{1}$$

where h_t is the return volatility at time t, which depends on the past returns via,

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}. \tag{2}$$

Here, ϵ_t is a sequence of independent and identically distributed random variables with $E\left[\epsilon_t^2\right] = 1$ and density function $f_\epsilon(\cdot)$. Note that, in order to ensure that h_t is positive for all time t, it is usually assumed that

$$\omega > 0, \quad \alpha_1, \ldots, \alpha_{p-1} \geq 0, \quad \alpha_p > 0, \quad \beta_1, \ldots, \beta_{q-1} \geq 0 \text{ and } \beta_q > 0.$$

Note that it is not assumed that $E\left[\epsilon_t\right] = 0$, which is a common assumption in the QMLE framework. Relaxing the condition $E\left[\epsilon_t\right] = 0$, will allow a more flexible structure on the innovation distribution. As a consequence the volatility h_t is interpreted as the conditional second moment of the returns instead of the conditional variance as is usually the case. The return process r_t following the model defined in Eqs. (1)

and (2) is strictly stationary if the following condition holds (see, Bougerol and Picard, 1992),

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{3}$$

A necessary and sufficient condition is also given in Bougerol and Picard (1992) but it is more complicated and difficult to check. On the other hand, when $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, the model defined in Eqs. (1) and (2) reduces to the IGARCH model of Engle and Bollerslev (1986).

The unconditional distribution of the returns, denoted by $f_r(\cdot)$, is determined by the distribution of the unobservable volatility h_t, denoted by $f_h(\cdot)$, and the distribution of the innovations, $f_\epsilon(\cdot)$, as follows,

$$f_r(r_t) = \int_0^\infty \frac{1}{\sqrt{h_t}} f_\epsilon(\epsilon_t) f_h(h_t) \, dh_t.$$

Some usual assumptions on $f_\epsilon(\cdot)$ includes the Gaussian, Student-t, Gaussian mixture, logistic, double exponential, Cauchy and generalized hyperbolic distributions, among others. The aim of this paper is to construct robust alternatives to the usual distributional assumptions on the innovations. In order to facilitate our proposed model, we rewrite the GARCH model defined in Eqs. (1) and (2) as follows,

$$r_t = \widetilde{h}_t^{1/2} \xi_t, \tag{4}$$

where $\widetilde{h}_t = h_t / \omega$ is a rescaled volatility given by,

$$\widetilde{h}_t = 1 + \sum_{i=1}^{p} \tilde{\alpha}_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \widetilde{h}_{t-j}, \tag{5}$$

where $\tilde{\alpha}_i = \alpha_i / \omega$ and $\xi_t = \omega^{1/2} \epsilon_t$ is a sequence of independent and identically distributed random variables with $E\left[\xi_t^2\right] = \omega$. We propose to assume that the rescaled innovation process ξ_t follows a location-scale mixture of Gaussian distributions, which is a broad class of distributions and includes all the distributions mentioned above. The density function of a location-scale mixture of Gaussian distributions (with respect to Lebesgue measure) is given by,

$$f_\xi(\xi_t|G) = \int \phi\left(\xi_t|\mu, \sigma^2\right) dG\left(\mu, \sigma^2\right), \tag{6}$$

where $\phi\left(\xi_t|\mu, \sigma^2\right)$ denotes the density function of the Gaussian distribution with mean μ and variance σ^2

and G is the location-scale mixing distribution. The key feature of the above semiparametric approach is the assumption that the location-scale mixing distribution G in the density function in Eq. (6) is unknown, and is modeled by a Dirichlet process (DP) prior, as will be described in Section 4, resulting in a DP mixture (DPM) model. Although new to the GARCH literature, DPM models have an extensive literature in Bayesian analysis and provide a broad and flexible class of distributions in many different settings, see, for instance, Ishwaran and Zarepour (2002), Basu and Chib (2003) and Ghosh, Basu and Tiwari (2009) and the references therein. In what follows, the model defined in Eqs. (4) and (5) where the innovations ξ_t are assumed to follow the density function in Eq. (6) and the location-scale mixing distribution G is modeled by a DP prior is called the Generalized Autoregressive Heteroscedastic Dirichlet Process Mixture Model (GARCH-DPM) model.

4   Bayesian inference for the GARCH-DPM model

This section describes how to perform Bayesian inference for the GARCH-DPM model where the location-scale mixing distribution G in Eq. (6) is modeled by a DP prior, leading to a DPM distribution for ξ_t. Inference for the GARCH-DPM model is developed using an MCMC algorithm which provides with samples from the joint posterior distribution of the model parameters. The algorithm is based on the one proposed in Papaspiliopoulos (2008) for DPM models, which is itself a combination of the procedures developed in Walker (2007) and Papaspiliopoulos and Roberts (2008).

Given a return series, $\boldsymbol{r} = \{r_1, \ldots, r_T\}$, it is assumed that the rescaled innovation process, ξ_t, follows a DPM distribution with concentration parameter ν and baseline probability measure G_0, which can be written hierarchically, for $t = 1, \ldots, T$, as,

$$
\begin{gathered}
\xi_t | \mu_t, \sigma_t^2 \sim N\left(\mu_t, \sigma_t^2\right) \\
\left(\mu_t, \sigma_t^2\right) | G \overset{iid}{\sim} G \\
G | \nu, \kappa, G_0 \sim DP\left(\nu, G_0\left(\cdot | \kappa\right)\right) \\
\left(\kappa, \nu\right) \sim \pi\left(\kappa\right) \pi\left(\nu\right),
\end{gathered}
\tag{7}
$$

where $\pi\left(\kappa\right)$ and $\pi\left(\nu\right)$ are the prior probabilities of the set of hyperparameters, κ, and the concentration parameter, ν, respectively.

Since the seminal work by Ferguson (1973), where the properties and theory about DP models were developed, several characterizations for DPM models have been given in the literature. For instance, Blackwell and MacQueen (1973) proposed a Pólya urn representation which was included within a Markov chain sampling by Escobar (1994) and Escobar and West (1995). Also, a Pólya urn structure was considered in the collapsed cluster sampling method and the "no-gaps" algorithm for nonconjugate DPM models, proposed respectively by MacEachern (1994) and MacEachern and Müller (1998). On the other hand, Sethuraman and Tiwari (1982) and Sethuraman (1994) proposed an alternative characterization of Dirichlet process mixtures in terms of a "stick-breaking" construction, which was furthermore extended by Ishwaran and Zarepour (2002) and Ishwaran and James (2001, 2003) and, more recently, by Walker (2007) and Papaspiliopoulos and Roberts (2008). Using this stick-breaking representation, the distribution of the auxiliary variable, ξ_t, introduced above, can be described hierarchically as ,

$$\xi_t \sim N(\mu_{z_t}, \sigma_{z_t}^2), \qquad \text{for } t = 1, \ldots, T$$
$$\Pr\left(z_t = s\right) = \rho_s, \qquad \text{for } s = 1, 2, \ldots$$
$$\left(\mu_s, \sigma_s^2\right) \sim G_0\left(\cdot|\kappa\right), \qquad \text{for } s = 1, 2, \ldots \tag{8}$$
$$\rho_1 = \eta_1, \qquad \rho_s = \left(1 - \eta_1\right) \ldots \left(1 - \eta_{s-1}\right) \eta_s$$
$$\eta_s \sim Be\left(1, \nu\right)$$

In particular, a conjugate model is assumed where $G_0\left(\cdot|\kappa\right)$ is a normal-gamma distribution $NG\left(m, \gamma, a, b\right)$, with parameters $\kappa = \left(m, \gamma, a, b\right)$, and is given by,

$$\mu_s|\sigma_s^2 \sim N\left(m, \frac{\sigma_s^2}{\gamma}\right),$$
$$\sigma_s^2 \sim IG\left(\frac{a}{2}, \frac{b}{2}\right),$$

where IG denotes the inverse gamma distribution. The concentration parameter, ν, and the set of hyperparameters, κ, will be considered fixed, although hyperprior distributions could be easily incorporated. For the illustration, it has been set $\nu = 1$, $a = 0.1$, $b = 0.1$, $m = 10^{-4}$ and $\gamma = 0.1$.

In order to complete the Bayesian formulation of the model, we also need to define prior distributions for the parameters $\tilde{\alpha}_i$ and β_j in the volatility equations (5). Uniform prior distributions are assumed for both

$\tilde{\alpha}_i$ and β_j, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, restricted to the stationary region,

$$\omega \sum_{i=1}^{p} \tilde{\alpha}_i + \sum_{j=1}^{q} \beta_j < 1. \tag{9}$$

Recall that $\omega = E\left[\xi_t^2\right]$ and then, noting that the DPM model described in Eq. (8) can be seen as an infinite mixture of Gaussian distributions given by,

$$f\left(\xi_t\right) = \sum_{s=1}^{\infty} \rho_s \phi\left(\xi_t \mid \mu_s, \sigma_s^2\right), \tag{10}$$

it is obtained that,

$$\omega = \sum_{s=1}^{\infty} \rho_s \left(\mu_s^2 + \sigma_s^2\right). \tag{11}$$

Finally, observe that as the stationary condition in Eq. (9) depends on ω, it is also necessary to impose a priori the condition in Eq. (9) on the DPM parameters, $(\rho_s, \mu_s, \sigma_s^2)$, $s = 1, 2, \ldots$, defined in Eq. (8).

Now, the aim is to construct an MCMC algorithm to sample from the joint posterior distribution of the model parameters, $\boldsymbol{\theta} = \left(\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}\right)$, where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots)$, $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \sigma_2^2, \ldots\right)$, $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$ and the missing data $\boldsymbol{z} = (z_1, \ldots, z_T)$. It would seem that infinite values should be sampled at each step of the MCMC algorithm. However, this will not be necessary following the ideas proposed in Walker (2007), which are based on slice sampling schemes (Damien, Wakefield and Walker, 1999).

Using these ideas, we introduce a latent variable u_t such that,

$$f\left(\xi_t, u_t\right) = \sum_{s=1}^{\infty} I\left(u_t < \rho_s\right) \phi\left(\xi_t \mid \mu_s, \sigma_s^2\right), \tag{12}$$

where I is the indicator function. Observe that by integrating over u_t, the marginal density in Eq. (10) is obtained. Also note that Eq. (12) can be written as,

$$f\left(\xi_t, u_t\right) = \sum_{s=1}^{\infty} \rho_s U\left(u_t \mid 0, \rho_s\right) \phi\left(\xi_t \mid \mu_s, \sigma_s^2\right),$$

and then, with probability ρ_s, ξ_t and u_t follow a normal and a uniform distribution, respectively. With the new set of missing data, $\boldsymbol{u} = (u_1, \ldots, u_T)$, the complete likelihood function for the rescaled innovation series,

$\boldsymbol{\xi} = (\xi_1, \dots, \xi_T)$, is,

$$f\left(\boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{z} \mid \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \propto \prod_{t=1}^{T} I\left(u_t < \rho_{z_t}\right) \phi\left(\xi_t \mid \mu_{z_t}, \sigma_{z_t}^2\right),$$

and consequently, the complete likelihood function for the observed time series, $\boldsymbol{r}$, is given by,

$$f\left(\boldsymbol{r}, \boldsymbol{u}, \boldsymbol{z} \mid \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}\right) \propto \prod_{t=1}^{T} I\left(u_t < \rho_{z_t}\right) \phi\left(r_t \mid \tilde{h}_t^{1/2} \mu_{z_t}, \tilde{h}_t \sigma_{z_t}^2\right).$$

Now, we can obtain the conditional posterior distributions of $\left(\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{u}, \boldsymbol{z}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}\right)$ as follows. Firstly, as suggested in Papaspiliopoulos (2008), the conditional posterior of $(\boldsymbol{\eta}, \boldsymbol{u})$ can be sampled jointly using,

$$f\left(\boldsymbol{\eta}, \boldsymbol{u} | \cdot\right) = f\left(\boldsymbol{u} \mid \boldsymbol{\eta}, \cdot\right) f\left(\boldsymbol{\eta} | \cdot\right).$$

Conditionally on $\boldsymbol{z}$, the posterior distribution of $(\boldsymbol{\eta}, \boldsymbol{u})$ is independent of $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}\right)$. In particular, it is easy to see that the conditional posterior distribution of u_t given $(\eta_1, \dots, \eta_{z_t})$ follows a uniform density on the interval $(0, \rho_{z_t})$, for $t = 1, \dots, T$, where $\rho_{z_t} = (1 - \eta_1) \dots (1 - \eta_{z_t-1}) \eta_{z_t}$. On the other hand, the marginal conditional posterior of η_s given $\boldsymbol{z}$ is,

$$\eta_s \sim Beta\left(n_s + 1, T - \sum_{l=1}^{s} n_l + \nu\right) \tag{13}$$

where,

$$n_s = \sum_{t=1}^{T} I\left(z_t = s\right)$$

is the number of observations assigned to the s-th mixture component.

Conditionally on $(\boldsymbol{z}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta})$, the posterior distribution of $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ is independent of $(\boldsymbol{\eta}, \boldsymbol{u})$, conjugate and given by,

$$\left(\mu_s, \sigma_s^2\right) \sim \left\{ \begin{array}{ll} NG\left(m, \gamma, a, b\right), & \text{if } z_t \neq s \text{ for any } t \in \{1, \dots, T\} \\ \\ NG\left(\tilde{m}_s, \tilde{\gamma}_s, \tilde{a}_s, \tilde{b}_s\right), & \text{if } z_t = s \text{ for some } t \in \{1, \dots, T\} \end{array} \right\} \tag{14}$$

where,

$$\tilde{m}_s = \frac{\gamma m + n_s \bar{\xi}_s}{\gamma + n_s}, \qquad \tilde{\gamma}_s = \gamma + n_s,$$

$$\tilde{a}_s = a + n_s, \qquad \tilde{b}_s = b + \sum_{t:z_t=s} \left(\xi_t - \bar{\xi}_s\right)^2 + \frac{\gamma n_s}{\gamma + n_s} \left(m - \bar{\xi}_s\right)^2$$

12

and,

$$\bar{\xi}_s = \frac{1}{n_s} \sum_{t:z_t=s} \xi_t.$$

Conditionally on $\left(\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{u}\right)$, the posterior distribution of $\boldsymbol{z}$ is independent of $(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta})$ and is given by,

$$\Pr\left(z_t = s\right) \propto I\left(s \in A_\rho\left(u_t\right)\right) \phi\left(\xi_t \mid \mu_s, \sigma_s^2\right) \tag{15}$$

where,

$$A_\rho\left(u_t\right) = \left\{s : \rho_s > u_t\right\}.$$

Observe that $A_\rho\left(u_t\right)$ is a finite set. This fact makes possible to sample from the posterior distribution of z_t, which would much more complicated without the use of the latent variables, u_t, see Walker (2007) for details. Then, we need to sample all the values of ρ_s which are larger than u_t. For this, Walker (2007) shows that it is enough to find the smallest s^* such that,

$$\sum_{s=1}^{s^*} \rho_s > 1 - u^*, \tag{16}$$

where $u^* = \min\{u_1, \ldots, u_T\}$.

Finally, the conditional posterior distribution of $\tilde{\alpha}_i$, for $i = 1, \ldots, p$, is independent of $(\boldsymbol{\eta}, \boldsymbol{u})$ and its kernel is given by,

$$k\left(\tilde{\alpha}_i \mid \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{z}, \boldsymbol{\beta}\right) = \prod_{t=1}^{T} \tilde{h}_t^{-1/2} \exp\left(-\frac{\left(r_t - \tilde{h}_t^{1/2} \mu_{z_t}\right)^2}{2\tilde{h}_t \sigma_{z_t}^2}\right), \tag{17}$$

restricted to the stationary region in Eq. (9). This posterior distribution is not of a standard form and then, random values cannot be not straightforwardly generated. However, it is possible to make use of the Griddy-Gibbs sampling algorithm, introduced by Ritter and Tanner (1992), which uses a numerical approximation of the cumulative distribution function. The Griddy-Gibbs sampler have been previously considered to make Bayesian inference for several financial time series models, see e.g. Bauwens and Lubrano (1998) and Ausin and Galeano (2007). The two main steps of the Griddy-Gibbs sampler are:

1. Approximate the following integral using a numerical integration method,

$$\Phi_g = \int_{x_1}^{x_g} k\left(x \mid \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{z}, \boldsymbol{\beta}\right) dx,$$

for $g = 1, \ldots, G$, where $x_1, \ldots, x_G$ is a grid of points in the domain of $\tilde{\alpha}_i$.

2. Generate $u \sim U(0, \Phi_G)$ and invert $\Phi(x)$ by numerical interpolation to obtain a draw of $\tilde{\alpha}_i$.

Clearly, the kernel of the conditional posterior distribution of β_j, for $j = 1, \ldots, p$, has the same expression as the given in Eq. (17) and analogously, the Griddy-Gibbs sampling algorithm can be used to sample from it.

Thus, the scheme of the MCMC algorithm would be as follows.

1. `Set an initial allocation` $z = \{z_1, \ldots, z_T\}$.

   `Iterate the following steps:`

2. `Update` η_s `by simulating from the beta distribution given in Eq. (13) for` $s = 1, \ldots, z^*$,
   `where` $z^* = \max\{z_t\}_{t=1}^T$.

3. `Update` u_t `by simulating from` $u_t \sim Uniform(0, \rho_{z_t})$ `for` $t = 1, \ldots, T$.

4. `Update` η_s `by simulating from` $\eta_t \sim Beta(1, \nu)$ `for` $s = z^* + 1, \ldots, s^*$, `where` s^* `is defined in`
   `Eq. (16)`.

5. `Update` (μ_s, σ_s^2) `by simulating from the normal-gamma distribution given in Eq. (14) for` $s = $
   $1, \ldots, s^*$.

6. `Update` z_t `by simulating from the discrete distribution given in Eq. (15) for` $t = 1, \ldots, T$.

7. `Update` $\tilde{\alpha}_i$ `by simulating from (17) for` $i = 1, \ldots, p$, `using the Griddy-Gibbs sampling.`

8. `Update` β_j `by simulating from (17) for` $j = 1, \ldots, p$, `using the Griddy-Gibbs sampling.`

Recall that the stationary condition in Eq. (9) have been imposed in the model parameters. This restriction can be incorporated in the algorithm by simply rejecting the simulated values in steps 1-5 when this condition is not verified. Observe that Eq. (9) depends on ω, which is theoretically known given $(\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, see Eq. (11). However, in practice, in order to evaluate this infinite sum, the following approximation can be used at every iteration of the MCMC algorithm,

$$\hat{\omega} = \sum_{s=1}^{z^*} \frac{n_s}{T} \left(\mu_s + \sigma_s^2 \right), \tag{18}$$

where $z^* = \max\{z_t\}_{t=1}^T$.

Another important issue in the algorithm is the choice of the grids to sample from the conditional posterior distributions of $\tilde{\alpha}_i$ and β_j. In the illustrations, an initial grid of 200 points in the interval $(0, 1)$ have been chosen for both α_i and β_i. This grid is modified in each step of the algorithm to a grid of points in the interval $(0, 1/\hat{\omega})$ for $\tilde{\alpha}_i$. Also, in order that the stationary condition in Eq. (9) holds, the kernel in Eq. (17) is only evaluated in those points of the grid where this condition is true for each parameter.

5 Deviance Information Criteria for model selection

In this section, the objective is to compare the proposed semiparametric model with other parametric approaches using the Deviance Information Criteria (DIC) proposed in Spiegelhalter, Best, Carlin and Van der Linde (2002). The smaller the DIC the better the model is. However, since the semiparametric models that we propose use mixture structures, the original definition of DIC is not applicable. In particular, we considered one of the modifications of the DIC recently proposed by Celeux, Forbes, Robert and Titterington (2006) for missing data models.

Usually, the DIC is defined by,

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D = -4 E_\theta \left[\log f \left(\boldsymbol{r} \mid \boldsymbol{\theta} \right) \mid \boldsymbol{r} \right] + 2 \log f \left(\boldsymbol{r} \mid \bar{\boldsymbol{\theta}} \right),$$

where $\overline{D(\boldsymbol{\theta})}$ denotes the posterior mean of the deviance $D(\boldsymbol{\theta}) = -2 \log f(\boldsymbol{r} \mid \boldsymbol{\theta})$ and p_D is the effective dimension given by $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, where $\bar{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ given the data $\boldsymbol{r}$. The usual choice for $\bar{\boldsymbol{\theta}}$ is the posterior mean of $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta} \mid \boldsymbol{r}]$. However, as pointed out in Celeux et al. (2006), in mixture models, the parameters $\boldsymbol{\theta}$ are not always identifiable and the posterior mean can then be a very poor estimator. In fact, this choice will often lead to negative values of the effective dimension, p_D. Celeux et al. (2006) define eight different modifications of the DIC. Considering that the semiparametric model proposed here can be viewed as an infinite mixture of Gaussian distributions, the DIC_4 criterium is chosen, which is the "complete DIC" suggested in Celeux et al. (2006) for mixture models. It is given by,

$$\text{DIC}_4 = -4 E_{\boldsymbol{\theta}} \left\{ E_{\boldsymbol{z}, \boldsymbol{u}} \left[\log f \left(\boldsymbol{r}, \boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{r}, \boldsymbol{\theta} \right) \right] \mid \boldsymbol{r} \right\} + 2 E_{\boldsymbol{z}, \boldsymbol{u}} \left[\log f \left(\boldsymbol{r}, \boldsymbol{z}, \boldsymbol{u} \mid E \left[\boldsymbol{\theta} \mid \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{u} \right] \right) \mid \boldsymbol{r} \right].$$

For the proposed GARCH-DPM model, it is obtained that,

$$E_{\boldsymbol{z},\boldsymbol{u}}[\log f(\boldsymbol{r},\boldsymbol{z},\boldsymbol{u}\mid\boldsymbol{r},\boldsymbol{\theta})]=-4\sum_{t=1}^{T}\sum_{s:\rho_s>u_t}\Pr(z_t=s\mid\boldsymbol{r},\boldsymbol{\theta})\log\phi\left(r_t\mid\tilde{h}_t^{1/2}\mu_s,\tilde{h}_t\sigma_s^2\right)$$

with,

$$\Pr(z_t=s\mid\boldsymbol{r},\boldsymbol{\theta})=\frac{\phi\left(r_t\mid\tilde{h}_t^{1/2}\mu_s,\tilde{h}_t\sigma_s^2\right)}{\sum\limits_{s:\rho_s>u_t}\phi\left(r_t\mid\tilde{h}_t^{1/2}\mu_s,\tilde{h}_t\sigma_s^2\right)}.$$

Thus, given the MCMC output, the first term of DIC_4 can be approximated by,

$$-4E_{\boldsymbol{\theta}}\left\{E_{\boldsymbol{z},\boldsymbol{u}}[\log f(\boldsymbol{r},\boldsymbol{z},\boldsymbol{u}\mid\boldsymbol{r},\boldsymbol{\theta})]\mid\boldsymbol{r}\right\}\simeq$$

$$-\frac{4}{K}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s:\rho_s^{(k)}>u_t^{(k)}}\Pr\left(z_t=s\mid\boldsymbol{r},\boldsymbol{\theta}^{(k)}\right)\log\phi\left(r_t\mid\sqrt{\tilde{h}_t^{(k)}}\mu_s^{(k)},\tilde{h}_t^{(k)}\sigma_s^{2(k)}\right).$$

while the second term in DIC_4 can be approximated by,

$$2E_{\boldsymbol{z},\boldsymbol{u}}\left[\log f(\boldsymbol{r},\boldsymbol{z},\boldsymbol{u}\mid E[\boldsymbol{\theta}\mid\boldsymbol{r},\boldsymbol{z},\boldsymbol{u}])\mid\boldsymbol{r}\right]\simeq\frac{2}{K}\sum_{k=1}^{K}\sum_{t=1}^{T}\log\phi\left(r_t\mid\sqrt{\tilde{h}_t^{(k)}}\bar{\mu}_{z_t^{(k)}},\tilde{h}_t^{(k)}\bar{\sigma}_{z_t^{(k)}}^2\right),$$

where

$$\bar{\mu}_s^{(k)}=E_{\boldsymbol{\theta}}\left[\mu_s\mid\boldsymbol{r},\boldsymbol{z}^{(k)}\right]=\frac{\gamma m+n_s^{(k)}\bar{\xi}_s^{(k)}}{\gamma+n_s^{(k)}}$$

$$\bar{\sigma}_s^{2(k)}=E_{\boldsymbol{\theta}}\left[\sigma_s^2\mid\boldsymbol{r},\boldsymbol{z}^{(k)}\right]=\frac{b+\sum_{t:z_t^{(k)}=s}\left(\xi_t-\bar{\xi}_s^{(k)}\right)^2+\frac{\gamma n_s^{(k)}}{\gamma+n_s^{(k)}}\left(m-\bar{\xi}_s^{(k)}\right)^2}{a+n_s-1}$$

with,

$$n_s^{(k)}=\sum_{t=1}^{T}I\left(z_t^{(k)}=s\right),\qquad\bar{\xi}_s^{(k)}=\frac{1}{n_s^{(k)}}\sum_{t:z_t^{(k)}=s}\xi_t^{(k)},$$

for all s such that $\rho_s^{(k)}>u_t^{(k)}$.

6   Volatilities and VaR estimation

In financial modeling, estimation of in-sample volatilities and prediction of future volatilities is a key feature.
Given the MCMC output, it is easy to obtain a sample from the posterior distribution of each unobserved

in-sample volatility h_t, for $t = 1, \ldots, T$. For that, we only need to calculate the value of h_t for each draw of the MCMC sample, $\boldsymbol{\theta}^{(k)}$, which is denoted by $h_t^{(k)}$. Then, the posterior mean for each volatility is estimated by,

$$E\left[h_t \mid \boldsymbol{r}\right] \simeq \frac{1}{K} \sum_{k=1}^{K} h_t^{(k)}, \tag{19}$$

where K is the size of the MCMC sample. Analogously, the posterior median is estimated using the sample median of $h_t^{(1)}, \ldots, h_t^{(K)}$ and credible intervals can be obtained using the corresponding quantiles of this sample. Using the same procedure, a sample from the posterior distribution of the one-step ahead volatility h_{T+1} can be obtained, which provides with estimates of the posterior mean and median of h_{T+1} and credible intervals.

On the other hand, the predictive density of r_{T+1},

$$f\left(r_{T+1} \mid \boldsymbol{r}\right) = E\left[f\left(r_{T+1} \mid \boldsymbol{r}, \boldsymbol{\theta}\right) \mid \boldsymbol{r}\right], \tag{20}$$

can be estimated as the mean of the density functions obtained for all the draws of the MCMC sample. Although each of these densities is an infinite mixture of Gaussians, the following approximation can be used,

$$f\left(r_{T+1} \mid \boldsymbol{r}\right) \simeq \frac{1}{K} \sum_{k=1}^{K} \hat{f}^{(k)}\left(r_{T+1}\right),$$

where,

$$\hat{f}^{(k)}\left(r_{T+1}\right) = \sum_{s=1}^{z^{*(k)}} \frac{n_s^{(k)}}{T} \phi\left(r_{T+1} \mid \sqrt{\tilde{h}_{T+1}^{(k)}} \mu_s^{(k)}, \tilde{h}_{T+1}^{(k)} \sigma_s^{2(k)}\right), \tag{21}$$

see e.g. Section 2 of Müller, Rosner, De Iorio and MacEachern (1996).

The measurement and management of risks have become one of the most challenging tasks in financial modeling because of the existence of extreme price movements in financial markets. Several kind of financial risk measures have been proposed including the VaR, expected shortfall and the spectral measures of risk. However, VaR has become the most widely used measure of market risk maybe because it has been adopted by central banks for analyzing capital adequacy. Broadly speaking, VaR indicates the potential loss of an asset over a certain time horizon at a certain risk level associated with an unfavorable movement in market prices. From the statistical point of view, given a certain probability level π, the t-period VaR at risk level π of a return series is defined as the value such that the probability that the losses at time t exceed this value is π. In other words, the t-period VaR at risk level π of a return series is given by the π-quantile of

the distribution of the return at time t,

$$\pi = \Pr \left(r_t \leq \mathrm{VaR}_{\pi,t} \right). \tag{22}$$

Observe that calculation of the VaR in Eq. (22) strongly depends on the assumption made for the innovation distribution. Thus a misspecified distribution can give a wrong estimate of VaR. Our semiparametric approach, viz., GARCH-DPM, however, is robust to misspecified distributions and thus appropriate for robust VaR estimation. Additionally, our proposed methodology allows to calculate the predictive distributions of the t-period VaR at risk level π, which, as a by-product, provides with a measure of precision for VaR estimates via predictive intervals. More specifically, given the MCMC sample, the posterior mean of $\mathrm{VaR}_{\pi,t}$ can be approximated using,

$$E \left[\mathrm{VaR}_{\pi,t} \mid \boldsymbol{r} \right] \approx \frac{1}{K} \sum_{k=1}^{K} \mathrm{VaR}_{\pi,t}^{(k)}, \tag{23}$$

where $\mathrm{VaR}_{T+1}^{(k)}$ is the value of $\mathrm{VaR}_{\pi,t}$ at the k-th iteration of the MCMC algorithm, which is given by,

$$\mathrm{VaR}_{\pi,t}^{(k)} = \sqrt{h_t^{(k)}} \xi_{t,\pi}^{(k)},$$

where $\xi_{t,\pi}^{(k)}$ is the π-quantile of the estimated mixture,

$$\hat{f}^{(k)} \left(\xi_t \right) = \sum_{s=1}^{z^{*(k)}} \frac{n_s^{(k)}}{T} \phi \left(\xi_t \mid \mu_s^{(k)}, \sigma_s^{2(k)} \right),$$

which can be obtained easily using a numerical approximation procedure such that the Newton-Raphson method. The sample of $\mathrm{VaR}_{\pi,t}^{(k)}$, for $k = 1, \ldots, K$, also allows to estimate the predictive median and Bayesian confidence intervals using the corresponding quantiles of this sample.

7 Simulation study

In this section, the proposed methodology is illustrated using one of the many artificial time series that have been analyzed to examine the performance of the developed procedure. A time series of size $T = 3000$ is simulated from the model defined in Eqs. (1) and (2) with,

$$h_t = 0.01 + 0.15 r_{t-1}^2 + 0.8 h_{t-1},$$

and the follwoing mixture distribution is assumed for the innovation process,

$$\epsilon_t \sim 0.9N\left(0.1, 0.5\right) + 0.1N\left(-1, 4.41\right), \tag{24}$$

which implies that $E\left[\epsilon_t\right] = -0.01$ and $E\left[\epsilon_t^2\right] = 1$. Observe that this mixture distribution leads to a non-centered and slightly left-skewed innovation distribution, both of which are usually observed in financial time series.

The proposed MCMC algorithm described in Section 4 is then run for a GARCH$(1, 1)$ model using 20000 iterations. The initial 10000 iterations are discarded as burn-in iterations in order to obtain a sample of size $K = 10000$ from the joint posterior distribution of the model parameters. The MCMC chains present a good mixing performance and fast convergence, which is checked using the test proposed by Geweke (1992). Figure 3 shows the trace plots and histograms of the posterior samples of the main model parameters: α, β, ω and ψ, where ψ represents the expectation of ϵ_t. Observe that the posterior sample of α has been obtained by evaluating $\alpha^{(k)} = \omega^{(k)}\tilde{\alpha}^{(k)}$, for $k = 1, \ldots, K$, where $\omega^{(k)}$ is obtained using Eq. (18) for each MCMC iteration. Also, considering that $\psi = E\left[\epsilon_t\right] = \omega^{-1/2}E\left[\xi_t\right]$, the posterior sample for ψ is obtained by evaluating,

$$\psi^{(k)} = \frac{1}{\sqrt{\hat{\omega}^{(k)}}} \sum_{s=1}^{z^{*(k)}} \frac{n_s^{(k)}}{T} \mu_s^{(k)},$$

for $k = 1, \ldots, K$.

Table 2 shows the posterior means and standard deviations of the model parameters using the proposed semiparametric approach and compare these estimations with those obtained using a single Gaussian distribution for the innovation distribution, which is a usual choice in financial models. Observe that the parameter estimation results are apparently very similar. However, note that the Bayesian DIC criterion clearly selects the semiparametric model. This is because there are important differences in the assumption of the innovation distribution. These differences are illustrated in Figure 4, where the true innovation distribution, given in Eq. (24), and the estimated densities obtained with the semiparametric and the Gaussian model are shown. Observe that the semiparametric model captures the left-skewed shape of the true distribution with a long left tail, while the Gaussian model leads to a symmetric predictive distribution. These differences between the two estimated densities will be very important in terms of uncertainty measure and Bayesian prediction of quantities of interests such as volatilities and VaR, as shown next.

Figure 5 illustrates the true values and posterior medians of the last 50 in-sample volatilities of the series
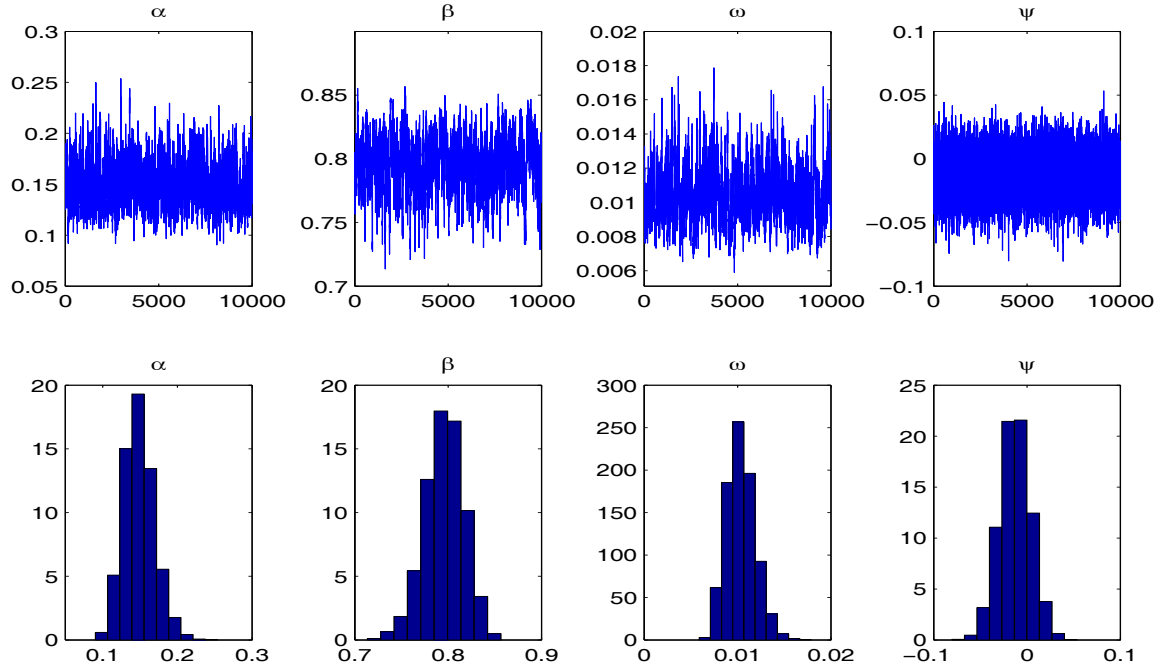
Figure 3: Trace plots and histograms of the posterior samples of the model parameters for the simulated data.

Table 2: Posterior means and standard deviations of the model parameters obtained for the simulated data using the proposed semiparametric model and a Gaussian distribution for the innovation process.

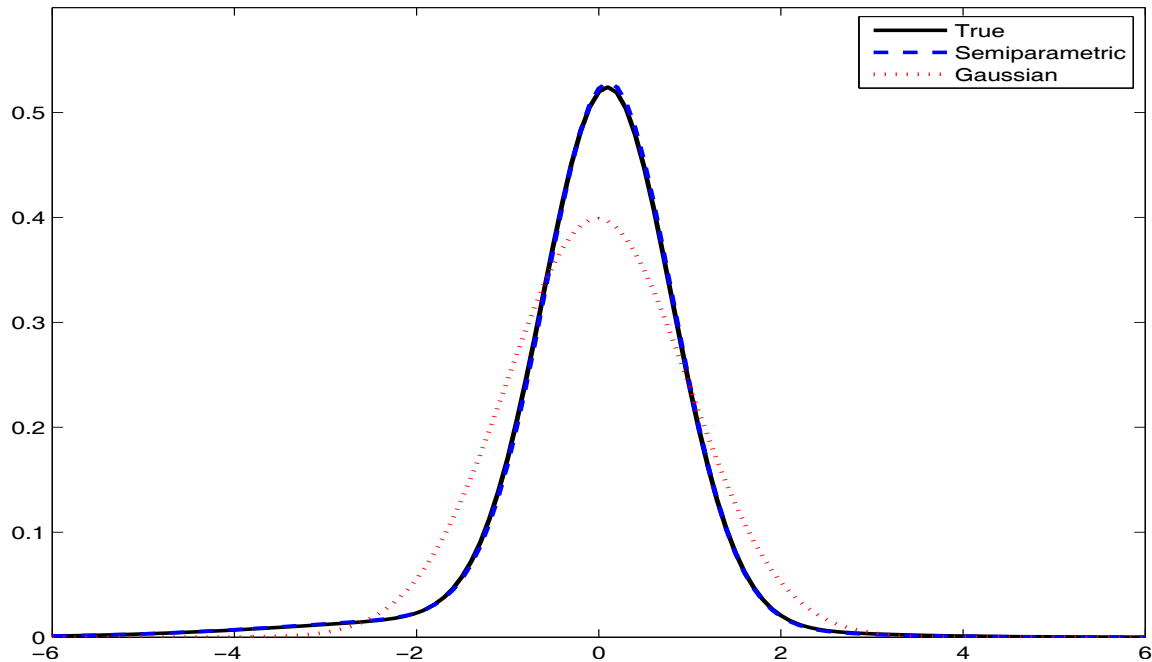|		Semiparametric	Gaussian
Parameter	True values	Posterior Mean posterior std.	Posterior Mean posterior std.
α	0.15	0.1489 0.0205	0.1456 0.0182
β	0.80	0.7957 0.0209	0.8070 0.0190
ω	0.01	0.0103 0.0015	0.0091 0.0012
ψ	-0.01	-0.0127 0.0167	-0.0135 0.0184
DIC		856.76	2337.09

Figure 4: Predictive densities of the innovation distribution for the simulated data.

with the 95% credible intervals obtained as described in Section 5, using the semiparametric and the Gaussian model. Observe that the Bayesian credible intervals always include the true generated volatilities for all time periods using the two models. However, observe that the estimations are closer to the true values using the semiparametric approach. Furthermore, the interval widths are slightly larger with the semiparametric model indicating more uncertainty in the estimations. Finally, observe that using the approach described in Section 5, it is also possible to obtain Bayesian predictions and credible intervals for the one-step ahead volatility, $h_{T+1} = h_{3001}$, with both the semiparametric and Gaussian model, as shown also in Figure 5, where they are compared with its true value.

Figure 6 shows the true values, posterior medians and 95% credible intervals for the t-period VaR at risk level $\pi = 0.01$, for $t = 2900, \ldots, 3001$, calculated as explained in Section 5, using the GARCH-DPM and Gaussian models. Observe that the Bayesian credible intervals obtained with the semiparametric approach always include the true values of VaR for all time periods, while the posterior medians of VaR obtained with the Gaussian model are far from the true values, which are never inside the credible intervals. Also, the length of the intervals is larger for the semiparametric model. Furthermore, the values of VaR are overestimated for all time periods. In particular, the Gaussian model predicts that the one-step ahead $VaR_{0.01,3001}$ will be larger than it turns out to be, indicating that the maximum loss of the asset will be smaller than it really
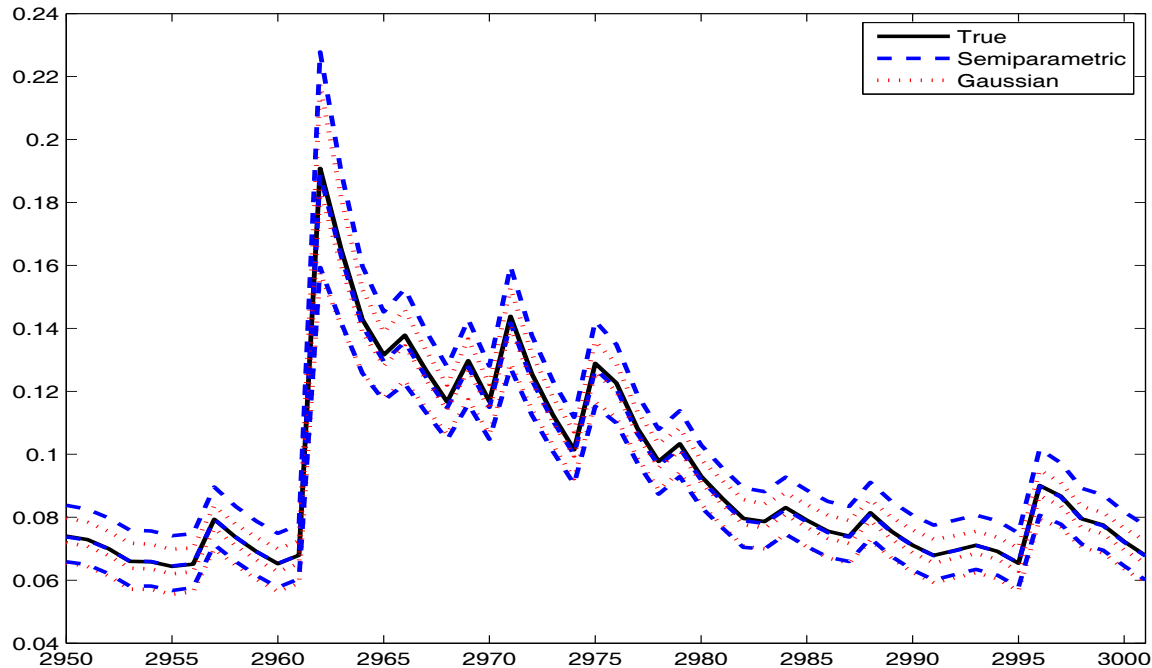
Figure 5: Posterior medians and 95% Bayesian confidence intervals for the conditional volatilities, h_t, for $t = 2950, \ldots, 3001$, for the simulated time series.

is.

8   Application

In this section the proposed methodology is applied to analyze the return time series of the BSE-30 and HSI indexes introduced in Section 2.

8.1   The Bombay Stock Exchange Index

In order to develop semiparametric Bayesian inference for the BSE-30 return series, the MCMC algorithm described in Section 3 is run for a GARCH(1,1) model using 20000 iterations and discarding the first 10000 as burnin iterations. The MCMC samples present good mixing performance (not shown to save space) and convergence is monitorized according to Geweke's statistic. Table 3 shows the posterior means and standard deviations for the main model parameters using the semiparametric approach and compares these estimations with those obtained using a simple Gaussian model for the innovation distribution. Observe that similar to the simulated data, there are no great differences in the parameter estimations. However, the DIC criterion
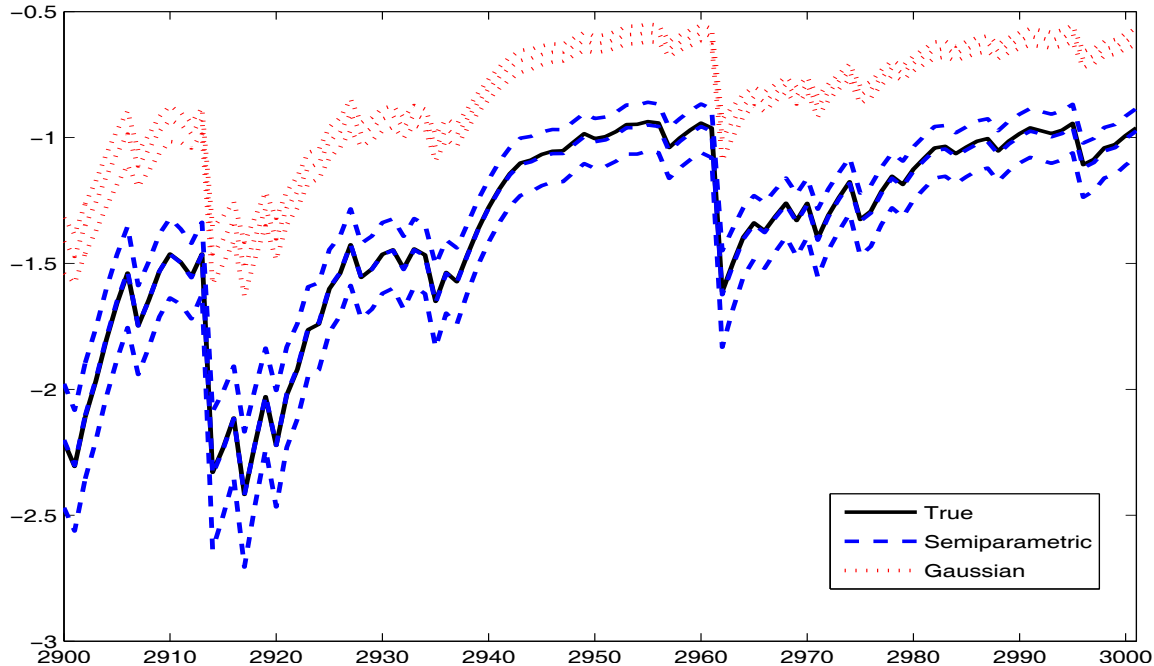
Figure 6: Posterior medians and 95% Bayesian confidence intervals for the $\mathrm{VaR}_{0.01,t}$, for $t = 2950, \ldots, 3001$, for the simulated time series.

clearly selects the semiparametric model. The main reasons for this choice are illustrated in Figure 7, where the predictive densities for the innovation distribution obtained with the semiparametric and the Gaussian model are shown. Observe that the left tail of the distribution is longer for the semiparametric model, while the Gaussian model leads to a symmetric estimation of the innovation density.

Figure 8 illustrates the posterior means of the volatility, h_t, obtained as described in (19), for the proposed semiparameric approach and for the Gaussian model. Note that both estimations are very similar and seem to describe properly the time-varying volatility of the BSE-30 returns shown in Figure 1. However, the Bayesian credible intervals are wider using the semipararametric approach for all time periods. This is illustrated in Table 4 where the posterior means, medians and 95% credible intervals for the one-step ahead volatility obtained with the semiparametric and the Gaussian model are shown.

Figure 9 (top) shows the BSE-30 returns together with the posterior means of the VaR at level $\pi = 0.01$ for all time periods using the proposed semiparametric approach. The proportion of times that the returns exceed the estimated VaR is 1.092%, which is very close to the theoretical value of 1%. It is important to note that when the same calculation is done but assuming a Gaussian model for the innovations, the proportion of times that the returns exceed the estimated VaR is 1.941%, which indicates that the estimated

Table 3: Posterior means and standard deviations of the model parameters using the proposed semiparametric model and a Gaussian distribution for the innovation process in the BSE-30 time series.

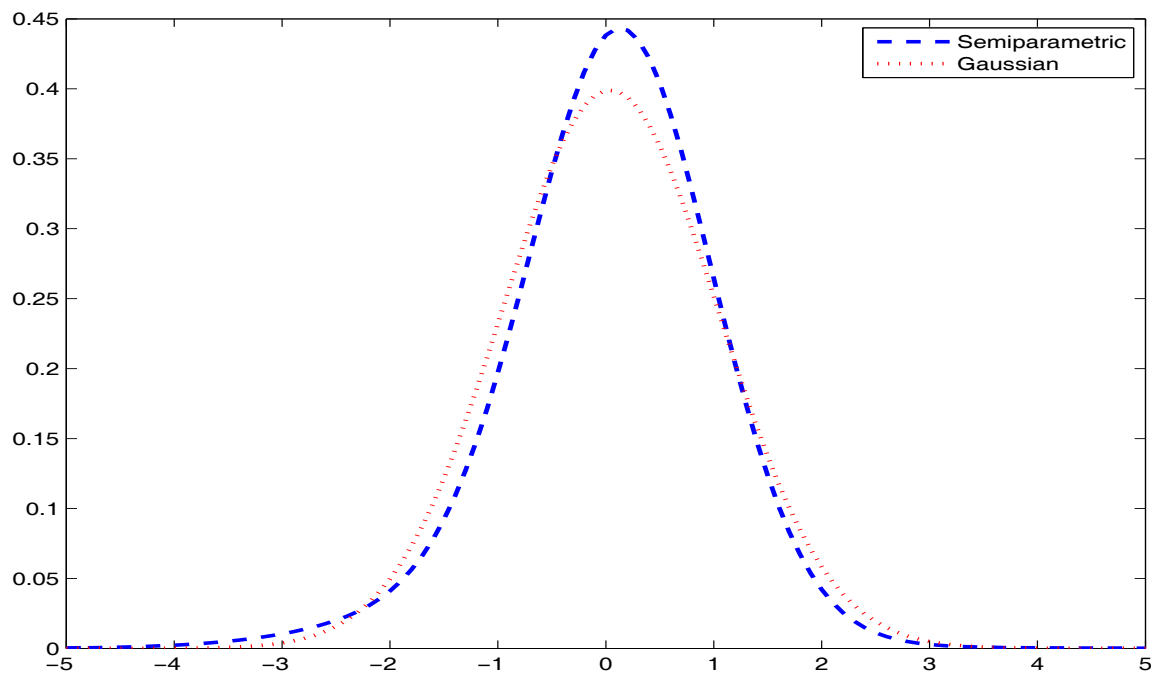|	Semiparametric	Gaussian
Parameter	Posterior Mean posterior std.	Posterior Mean posterior std.
α	0.1460 0.0203	0.1478 0.0202
β	0.8353 0.0200	0.8346 0.0213
ω	0.0819 0.0176	0.0761 0.0180
ψ	0.0401 0.0187	0.0404 0.0202
DIC	8595.38	9132.65

Figure 7: Predictive densities of the innovation distribution for the BSE-30 time series data.

Table 4: Posterior means, medians and 95% credible intervals for the one-step ahead volatility, h_{T+1}, assuming a semiparametric and a Gaussian model for the innovation process in the BSE-30 time series.

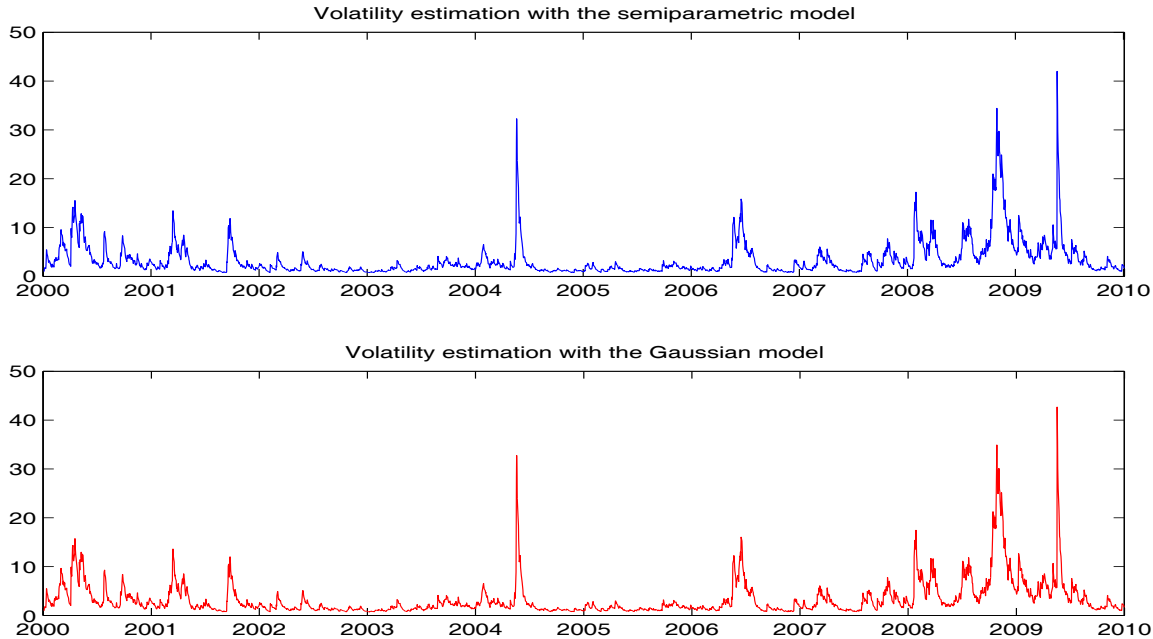| h_{T+1}	Posterior Mean	Posterior Median	95% Interval	Interval width
Semiparametric	1.6932	1.6873	(1.5430, 1.8760)	0.3330
Gaussian	1.6621	1.6637	(1.5167, 1.8010)	0.2843

Figure 8: Posterior means of the volatility, h_t, for the BSE-30 time series data assuming the proposed semiparametric approach (top) and a Gaussian model (bottom) for the innovation distribution.

maximum loss is exceeded considerably more than 1% of times. Furthermore, as for the simulated data, the 95% credible intervals for $\text{VaR}_{0.01,t}$ assuming the Gaussian model do not overlap those obtained with the semiparametric approach, which are always wider, and the posterior medians are always larger under the Gaussian model. This is illustrated in Figure 9 (bottom) where the Bayesian confidence intervals for the VaR at 1% level obtained with the two approaches are shown for the last three months of the time series. Consequently, all these results lead us to think that the Gaussian model is underestimating the maximum loss for all time periods.

8.2   The Hang Seng Index

For the HSI time series data, the MCMC algorithm described in Section 3 is run for a GARCH(1,1) model assuming a semiparametric Bayesian model for the innovation distribution. As for the previous data sets, a standard Bayesian inference assuming a Gaussian innovation distribution is also developed in order to examine the benefits of the proposed semiparametric approach. In both cases, 10000 burnin iterations are used followed by another 10000 iterations "in equilibrium". The traces of the MCMC samples present good mixing performance (not shown to save space) and convergence is checked using Geweke's statistic. Table 5
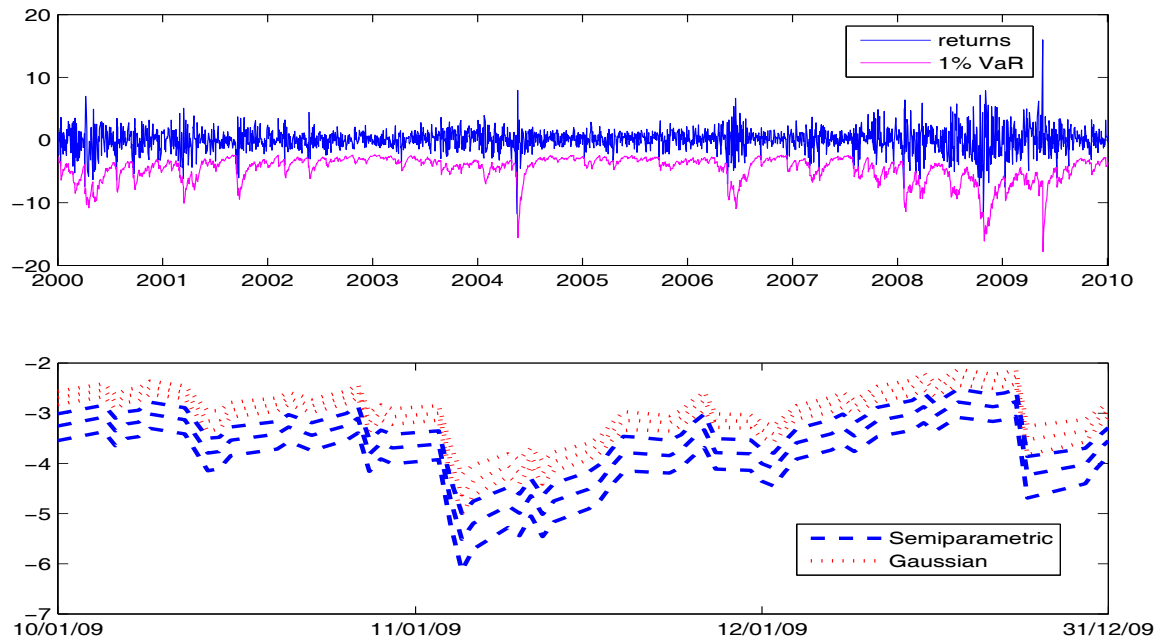
Figure 9: BSE-30 returns and posterior means of the $\text{VaR}_{0.01,t}$ obtained with the proposed semiparametric approach (top) and comparison of the posterior medians and 95% credible intervals of the $\text{VaR}_{0.01,t}$ for the last three months of the BSE-30 time series data using the semiparametric and Gaussian model (bottom).

shows the posterior means and standard deviations obtained with the two procedures. Observe that in this case there are more differences between the parameter estimation results obtained with the two approaches. Nevertheless, again for these data, the DIC criterion selects the semiparametric model. Figure 10 illustrates the estimated densities of the innovation distribution obtained with both approaches. Note that in this case the two predictive densities are symmetric, but the shape of the distributions are rather different. In particular, although it is hardly seen in the graph, we have observed that the tails of the distribution are slightly longer for the semiparametric model.

Figure 11 (top) shows the posterior means of the volatilities, h_t, using the proposed semiparametric approach. Observe that it seems to describe properly the time-varying volatility observed in the HSI returns shown in Figure 2. When these estimators are compared to those obtained using the Gaussian model, it is observed that they are similar, but there are more differences than those observed for the previous data sets. This is illustrated also in Figure 11 (bottom), where the posterior medians and credible intervals of the estimated volatilities with the semiparametric and the Gaussian approach are shown. Note that the posterior medians are slightly different and that, as in the previous examples, the credible intervals are wider for the semiparametric model.

Table 5: Posterior means and standard deviations of the model parameters using the proposed semiparametric model and a Gaussian distribution for the innovation process in the HSI time series.

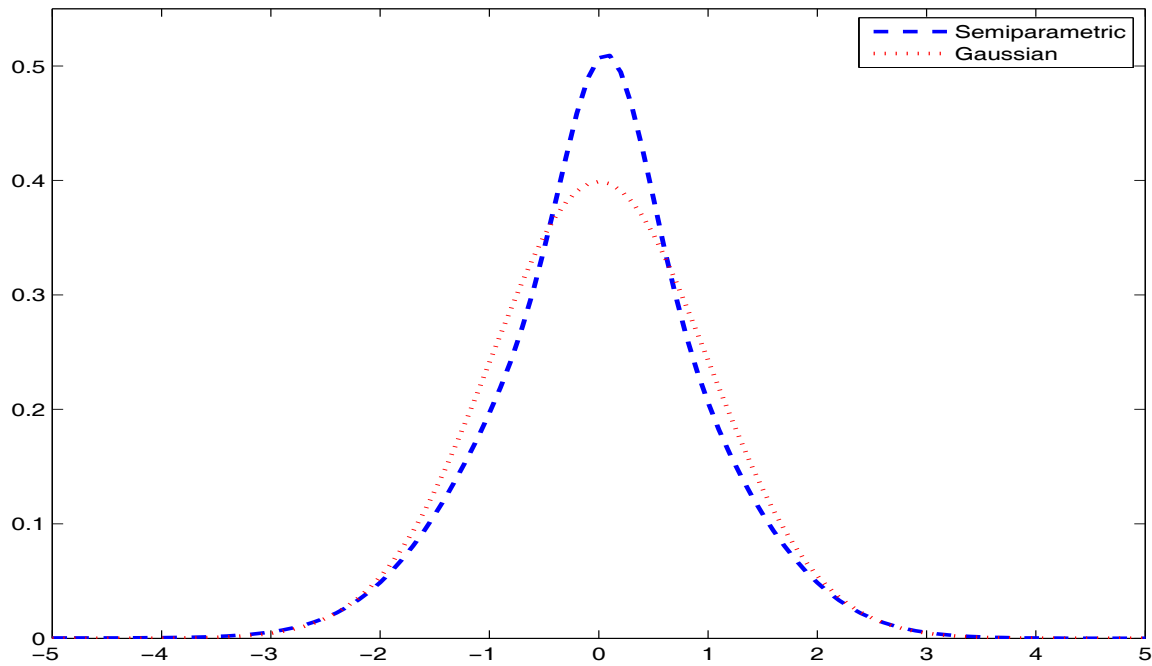| Parameter	Semiparametric Posterior Mean posterior std.	Gaussian Posterior Mean posterior std.
α	0.0895 0.0165	0.1321 0.0180
β	0.9009 0.0168	0.8534 0.0171
ω	0.0373 0.0091	0.0558 0.0110
ψ	0.0055 0.0195	0.0042 0.0199
DIC	8083.85	8811.31

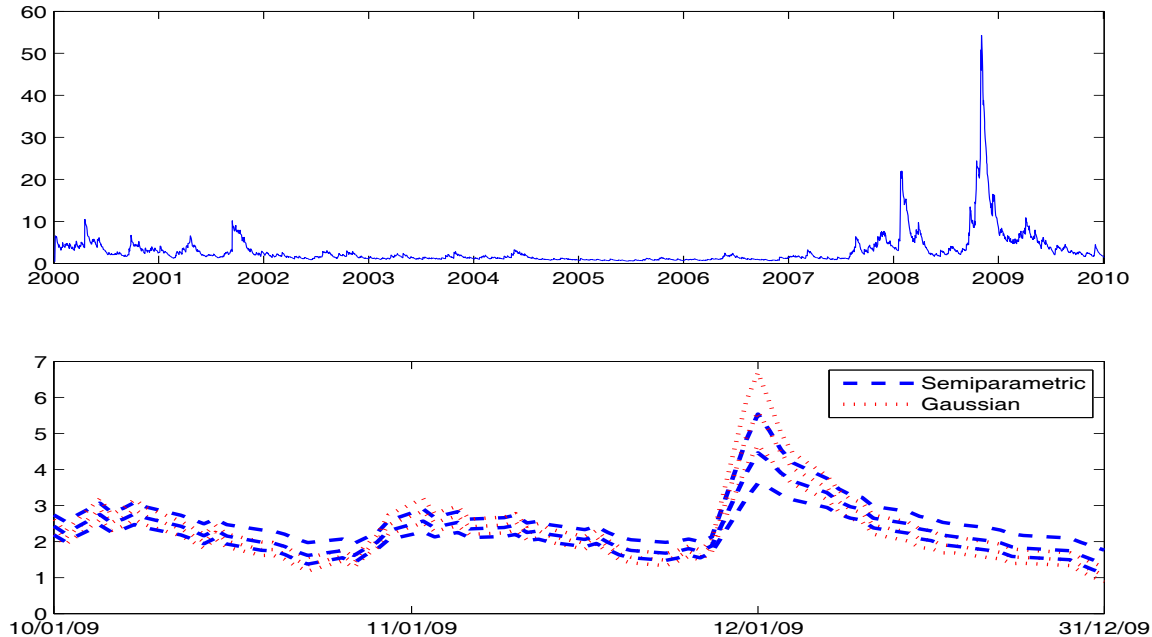Figure 10: Predictive densities of the innovation distribution for the HSI time series data.

Figure 11: Posterior means f the volatility, h_t, for the HSI time series data assuming the proposed semiparametric approach (top) and comparison of the posterior medians and 95% credible intervals of h_t for the last three months of the HSI time series data using the semiparametric and Gaussian model (bottom).

Finally, the $\text{VaR}_{t,\pi}$ is estimated for all time periods and for $\pi = 0.1$, 0.05 and 0.01 using the semiparametric and the Gaussian approach. Table 6 shows the proportion of times that the HSI returns exceed the estimated VaR in each case. It is also shown the difference in absolute value between this estimation and the theoretical value π. Observe that for $\pi = 0.1$ and 0.01, these differences are considerably smaller using the semiparametric approach and then, the VaR estimation is more accurate. For $\pi = 0.05$ the two differences are close indicating a similar VaR estimation. Then it seems that our proposed semiparametric GARCH-DPM model provides better estimations of VaR than those obtained with the Gaussian assumption.

Table 6: Proportion of times that the HSI returns exceed the estimated $\text{VaR}_{t,\pi}$ at different levels of π and differences in absolute value between these estimations and the true value π.

	$\pi = 0.1$	$\pi = 0.05$	$\pi = 0.01$
Semiparametric	0.1005 5.22×10^{-4}	0.0539 0.0039	0.0092 7.52×10^{-4}
Gaussian	0.0949 0.0051	0.0531 0.0031	0.0117 0.0017

9　Conclusions

In this article, a semiparametric Bayesian approach has been developed to make inference and prediction for univariate GARCH models. A location-scale mixture Gaussian model has been proposed for the innovation distribution using a Dirichlet process prior for the mixing distribution. An MCMC algorithm based on the combination of retrospective, slice and Griddy-Gibbs sampling have been constructed to obtain samples from posterior distributions of quantities of interest. It has been described how to compute the DIC criterion for this mixture data model which have been used for model selection among the proposed semiparametric approach and the usual Gaussian distribution. Furthermore, Bayesian prediction for volatilities and VaR has been carried out, obtaining Bayesian confidence intervals which are much more informative than simple point estimations.

The proposed methodology have been illustrated using a simulated data and two real financial times series: the Bombay Stock Exchange Index and the Hang Seng Index. In order to analyze the benefits of the proposed semiparametric approach, the obtained results has been compared to those obtained using a Gaussian model for the innovation distribution. The parameter estimation results are in general rather similar. However, there are important differences in the uncertainty measure of volatilities since, in general, credible intervals are wider using the semiparametric approach. More importantly, there are strong differences in the VaR estimation results. The Gaussian model usually underestimate the maximum loss of an asset due to an inadequate estimation of the tail of the distribution, while the semiparametric model seems to describe properly the asymmetric and long-tailed shape of the innovation distribution. Besides, Bayesian confidence intervals of VaR are wider with the semiparametric approach and the proportion of times that the VaR is exceed is in general closer to the true theoretical level.

It would be very interesting to generalize the proposed approach to the multivariate setting. An adequate description of the multivariate innovation distribution is essential in multivariate GARCH models. For example, it is well known that the VaR of a portfolio of assets strongly depends on the multiviate distribution of the innovation process. The theory and implementation of these issues are currently under research.

References

Antoniak, C. (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Annals of Statistics, 2, 1152–1174.

Ardia, D. (2008) Financial Risk management with Bayesian Estimation of GARCH Models: Theory and Applications. Springer

Ausin, M. C. and Galeano, P. (2007) Bayesian estimation of the Gaussian mixture GARCH model. Computational Statistics and Data Analysis, 51, 2636-2652.

Bai, X., Rusell, J. R., and Tiao, G. C. (2003) Kurtosis of GARCH and stochastic volatility models with non-normal innovations. Journal of Econometrics, 114, 349-360.

Basu, S. and Chib, S. (2003) Marginal Likelihood and Bayes Factors for Dirichlet Process Mixture Models. Journal of the American Statistical Association, 98, 224-235.

Basu, S. and Mukhopadhyay, S. (2000) Binary response regression with normal scale mixture links. In: D. K. Dey, S. Gosh and B. Mallick (Eds.), Generalized Linear Models: A Bayesian Perspective, Marcel Dekker, New York, 231-241.

Bauwens, L. and Lubrano, M. (1998) Bayesian inference on GARCH models using Gibbs sampler. Econometrics Journal, 1, 23-46.

Bauwens, L. and Lubrano, M.(2002) Bayesian option pricing using asymmetric GARCH models. Journal of Financial Econometrics, 9, 321-342.

Blackwell, D. and MacQueen, J. B. (1973) Ferguson distributions via Pólya Urn schemes. Annals of Statistics, 1, 353-355.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 51, 307-327.

Bollerslev, T. (1987) A conditional heteroscedastic time series model for speculative prices and rates of return. Review of Economics and Statistics, 69, 542-547.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992) ARCH modeling in Finance-A review of the theory and empirical evidence. Journal of Econometrics, 52, 5-59.

Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some non-negative time series. Journal of Econometrics, 52, 115-127.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006) Deviance Information Criteria for missing data models. Bayesian Analysis, 1, 651-674.

Damien, P., Wakefield, J. and Walker, S. (1999) Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. Journal of the Royal Statistics Society, Series B, 61, 331-344

Engle, R. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the U.K. inflation. Econometrica, 50, 987-1008.

Escobar, M. D. (1994) Estimating normal means with a dirichlet process prior. Journal of the American Statistical Association, 89, 268–277.

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

Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. Annals of Statistics, 1, 209–230.

Fernández, C. and Steel, M. F. J. (1998) On Bayesian modelling of fat tails and skewness. Journal of the American Statistical Association, 93, 359-371.

Galeano, P. and Ausin, M. C. (2009) The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter estimation, Value at Risk calculation and portfolio selection. Journal of Business and Economic Statistics, in press.

Gallant, A. R. and Tauchen, G. (1989) Semi-nonparametric estimation of conditionally constrained heterogeneous processes: asset prices application. Econometrica, 57, 1091-1120.

Gelfand, A. E., Dey, D. K. and Chang, H. (1992) Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In Bayesian Statistics 4, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford: Oxford University Press, 147-159.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol 4. Clarendon Press, Oxford.

Ghosh, P., Basu, S. and Tiwari, R. M. (2009) Bayesian analysis of cancer rates from SEER program using parametric and semiparametric joinpoint regression models. Journal of the American Statistical Association, 104, 439–452.

Glosten, L., Jaganathan, R. and Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance, 48, 1779–1801.

Hansen, B. E. (1994) Autoregressive conditional density estimation. International Economic Review, 35, 705-730.

Ishwaran, H. and James, L. (2001). Gibbs sampling methods for stick-breaking priors. Journal of the American Statististical Association, 96, 161-73.

Ishwaran, H. and James, L. (2003). Some further developments for stick-breaking priors: finite and infinite clustering and classification. Sankhya A 65, 577-92.

Ishwaran, H. and Zarepour, M. (2002) Exact and Approximate Sum Representations for the Dirichlet Process. Canadian Journal of Statistics, 30, 269-283.

Kaufmann, S. and Frühwirth-Schnatter, S. (2002) Bayesian Analysis of Switching ARCH Models. Journal

of Time Series Analysis, 23, 425-458.

Lee, Y. and Tse, T. (1991) Term structure of interest rates in the Singapore Asian Dollar market. Journal of Applied Econometrics, 6, 143-152.

MacEachern, S. N. (1994) Estimating normal means with a conjugate style Dirichlet process prior. Communications in Statistics Simulation and Computation, 23, 727-741.

MacEachern, S. N. and Müller, P. (1998) Estimating mixtures of Dirichlet process models. Journal of Computational and Graphical Statistics, 7, 223-238.

Müller, P., Rosner, G. L., De Iorio, M. and MacEachern (1996) Monte Carlo posterior integration in GARCH models. Sankhya, 60, 127-144.

Müller, P., Pole, A. (1998) Monte Carlo posterior integration in GARCH models. Sankhya, 60, 127-144.

Nakatsuma, T. (2000) Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. Journal of Econometrics, 95, 57-69.

Neal, R. (2000) Markov chain sampling: Methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics, 9, 249-265.

Nelson, D.B. (1991) Conditional heteroscedasticity in asset pricing: a new approach. Econometrica, 59, 347-370.

Papaspiliopoulos, O. (2008) A note on posterior sampling from Dirichlet mixture models. Working paper, Universidad Pompeu Fabra.

Papaspiliopoulos, O. and Roberts, G. O. (2008) Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. Biometrika, 95, 169-186.

Premaratne, G. and Bera, A. K. (2000) Modeling asymmetry and excess kurtosis in stock return data. Working paper, University of Illinois.

Ritter, C. and Tanner, M.A. (1992) Facilitating the Gibbs sampler: the Gibss stopper and the Griddy–Gibbs sampler. Journal of the American Statististical Association 87, 861-868.

Sethuraman, J. (1994) A constructive definition of Dirichlet priors. Statistica Sinica, 4, 639-650.

Sethuraman, J. and Tiwari, R. C. (1982) Converge of Dirichlet measure and the interpretation of their parameters. In Statistical Decisions Theory and Related Topics III (Vol. 2), eds. S. Gupta and J. O. Berger, Academic Press, 305-315.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, B, 64, 583-639.

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, 203–226. Amsterdam: North-Holland.

Taylor, S. J. (2005) Asset Price Dynamics, Volatility, and Prediction. Princeton University Press.

Tsay, R. S. (2005) Analysis of financial time series, 2nd edition. John Wiley & Sons, Inc.

Vrontos, I. D., Dellaportas, P. and Politis, D.N. (2000) Full Bayesian inference for GARCH and EGARCH models. Journal of Business and Economic Statistics, 18, 187-198.

Walker, S. G. (2007) Sampling for Dirichlet mixture model with slices. Communications in Statistics-Simulation and Computation, 36, 45-54.