Modelling nonlinearities and heavy tails via threshold normal mixture GARCH models

D. Giannikis, I.D. Vrontos, P. Dellaportas*

Department of Statistics, Athens University of Economics and Business, Patission 76, 10434 Athens, Greece

Received 27 September 2006; received in revised form 17 April 2007; accepted 7 May 2007
Available online 18 May 2007

Abstract

A new class of flexible threshold normal mixture GARCH models is proposed for the analysis and modelling of the stylized facts appeared in many financial time series. A Bayesian stochastic method is developed and presented for the analysis of the proposed model allowing for automatic model determination and estimation of the thresholds and their unknown number. A computationally feasible algorithm that explores the posterior distribution of the threshold models is designed using Markov chain Monte Carlo stochastic search methods. A simulation study is conducted to assess the performance of the proposed method, and an empirical application of the proposed model is illustrated using real data.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Autoregressive conditional heteroscedasticity; Bayesian inference; Markov chain Monte Carlo; Stochastic search; Value at Risk

1. Introduction

Many financial series exhibit features such as time varying volatility, heavy tailed distributions, large kurtosis, nonlinearities and extreme events. Modelling these stylized facts has attracted a lot of attention in recent years. Engle (1982) introduced the autoregressive conditional heteroskedastic (ARCH) model and Bollerslev (1986) extended this model to the generalized ARCH to capture the volatility clustering phenomenon. However, these models under the conditional normality assumption for the return series are not capable to explain sufficiently the kurtosis of financial series. To circumvent this problem, several heavy tailed conditional distributions have been proposed in the literature. Bollerslev (1987) proposed modelling returns by using the Student-$t$ density, while Nelson (1991) considered the generalized error distribution. To better account for heavy tails and large kurtosis Bai et al. (2003) proposed a mixture of two normal distributions for the analysis of stock and exchange rate returns. Recent empirical analysis using different types of mixture normal GARCH models indicate that this class of models provide a better fit to the data than many other GARCH-type models; see, for example, Alexander and Lazar (2004, 2006) and Haas et al. (2004).

On the other hand, it has often been found that the linear time series models are restrictive and usually leave certain aspects of financial and economic data unexplained. A popular class of nonlinear time series models that have been proposed in the literature is that of threshold autoregressive models (Tong, 1978; Tong and Lim, 1980). These
models can be used to model the nonlinearities in the conditional mean of a time series. Tong (1990) provides a description/discussion of many commonly used nonlinear models and suggests a threshold model with a changing conditional variance. Li and Li (1996) extended the threshold model to the so-called double threshold ARCH model, which can handle the situation where both the conditional mean and the conditional variance are piecewise linear, given previous information. These models allow the dynamics of financial series to be different (asymmetric) in different regimes/partitions of the data.

Other classes of nonlinear models have recently been proposed in the literature. One class of models is that of mixture autoregressive (MAR) models introduced by Le et al. (1996) and Wong and Li (2000). These models can be defined by specifying the conditional distribution of the observed time series as a mixture of normal distributions where each component follows a Gaussian autoregressive (AR) model. Wong and Li (2001) extended the class of MAR models to allow for heteroscedastic features, and proposed the class of MAR conditional heteroscedastic models. The relation/connection of MAR models to the traditional threshold AR models has been presented by Lanne and Saikkonen (2003) who extended the model allowing for GARCH-type conditional heteroskedasticity. One other flexible class of nonlinear time series models, called endogenous delay threshold autoregressive models (EDTAR) was proposed by Koop and Potter (2003), who found empirical evidence in favour of their model over traditional threshold models. They developed Bayesian methods for the analysis of EDTAR models based on Markov chain Monte Carlo (MCMC) techniques. They argued (see also, Koop and Potter, 1999a, b; So and Chen, 2003; Yu et al., 2006) that Bayesian methods are particularly attractive for modelling nonlinear time series because they account for parameter and model uncertainty.

In this study, we propose a new class of flexible threshold models that can be used to better capture the stylized facts that appear in many financial time series. We relax the assumption of the fixed number of regimes/partitions allowing for automatic model determination and estimation of the number and the values of thresholds. Within each regime/partition we describe the dynamics of the observed data by using a mixture of two normal distributions where each component follows a Gaussian autoregressive model with different variances. We also consider GARCH-type conditional heteroskedasticity that allows for time varying variances. Note that the class of threshold models we propose is general and can be used for any mixture GARCH specification that has been proposed in the literature. In particular, we extend two mixture GARCH models; the model proposed by Bai et al. (2003) that better explains the leptokurtosis observed in financial series, and the model introduced by Haas et al. (2004) which allows the conditional variances to be time varying in each component. Therefore, the threshold normal mixture (TNM) specification we introduce allows also for appropriate modelling of the nonlinearities that are often present in financial series.

We address the general problem of Bayesian inference, model selection and volatility prediction. We demonstrate that adoption of the Bayesian framework can be advantageous in terms of generality, flexibility and accuracy. It takes account of model uncertainty by using Bayesian model averaging (BMA). We develop a Bayesian approach for finding threshold structured normal mixture (NM)-GARCH models. Simulation of the posterior distribution of the proposed threshold model is achieved by designing an MCMC algorithm that visits (jumps between) a variety of models. It can be seen as a Bayesian-motivated stochastic search algorithm that produces threshold models together with their posterior probabilities. In this way the proposed Bayesian methodology explicitly accounts for model uncertainty; for a given set of threshold specifications inference is based on their posterior probabilities, thus avoiding the usual approach that considers the models separately and chooses the best model via significance tests.

The remainder of the paper is organized as follows. The Bayesian threshold structured NM-GARCH models and the prior specification are introduced in Section 2. Bayesian inference under model uncertainty and the proposed MCMC stochastic search algorithm are presented in Section 3. In Section 4 we perform a simulation study and report the results, while in Section 5 we illustrate the above methodology in real financial time series data. We conclude in Section 6 with a brief discussion.

2. TNM-GARCH models

Throughout the paper we consider having observed data of the form \( y = (y_1, y_2, \ldots, y_N)' \). The threshold structured NM-GARCH model has two main components: \( k \) partitions (\( k - 1 \) thresholds) and a specification of the NM-GARCH for every partition. Below, we present the threshold structure and the GARCH model specification, define the model parameters and specify the prior used in this study.
2.1. The class of threshold models

The threshold model we propose can be uniquely defined by the $k - 1$ thresholds which separate the return series $y$ into $k$ partitions $P_j$, $j = 1, \ldots, k$. The threshold structured NM-GARCH model parameterizes every partition by a mixture of normal distributions and allows the conditional variances to be time varying by following a GARCH specification. The number of partitions present in the model determines its stochastic structure complexity. A change in the number of the partitions $k$ leads to a change of the parameter space dimension. Inference is carried out assuming that the true model which is specified by the number and the threshold values is unknown. The parameter vector of a threshold model $T$ with $k$ partitions $(k - 1$ thresholds) is denoted by $\theta_T^{(k)} = (t_1, \ldots, t_{k-1})$, where $t_i, i = 1, \ldots, k - 1$ denote the threshold values of the model. Therefore, the threshold structure is determined uniquely by the number of partitions $k$ and the parameter vector $\theta_T^{(k)}$ of a model $T$ (with $k$ partitions). We will denote a specific model $T$ with $k$ partitions and threshold parameters $\theta_T^{(k)}$ by $T_{k, \theta}$. Bayesian inference about $k$ and $\theta_T^{(k)}$ is based on the joint posterior distribution $p(k, \theta_T^{(k)} | y)$ which can be written as

$$p(k, \theta_T^{(k)} | y) = p(k | y) p(\theta_T^{(k)} | k, y),$$

where $y$ is the observed data. We generate a sample from the posterior distribution of $(k, \theta_T^{(k)})$ using an MCMC algorithm that visits (jumps between) models of different dimensionality.

Due to the complexity of the NM-GARCH models we need a sufficient number of observations within each partition in order to estimate the model parameters. For this reason we allow the threshold parameters, $t_1, \ldots, t_{k-1}$, to take values from a prespecified finite grid. In the applications of this paper this grid consists of a set of empirical percentiles of the data at hand. This is a standard approach for this type of problems which involve complex models; see, for example, Audrino and Buhlmann (2001).

2.2. The NM-GARCH specification

A model with $k - 1$ thresholds assumes that there are $k$ partitions in the observation state. Given a specific partition, we describe the dynamics of $y_t$ in that partition by using a mixture of normal distributions where the conditional variances are specified by a GARCH formulation. In particular, we use two well-established NM-GARCH specifications, the NM model introduced by Bai et al. (2003) and the mixed normal conditional heteroscedasticity (MNCH) model of Haas et al. (2004).

2.2.1. The TNM-GARCH model

The NM-GARCH model proposed by Bai et al. (2003) and further analysed by Ausin and Galeano (2007) can be written in the form

$$y_t = m_t + n_t, \quad t = 1, \ldots, N,$$

where $m_t$ is the conditional mean of the process, $n_t$ is the error term given by $\sqrt{h_t} \varepsilon_t$ with $h_t$ being the conditional variance of $y_t$ given the previous information $\Phi_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, and $\varepsilon_t$ being a zero mean and variance one process. Bai et al. (2003) and Ausin and Galeano (2007) used a constant mean for the observation process ($m_t = 0$ and $\mu$, respectively) and a mixture of two normal distributions for innovations $\varepsilon_t$ together with a GARCH specification for the conditional variance, $h_t$. In particular, they proposed the following mixture of distributions:

$$\varepsilon_t \sim \begin{cases} N(0, \sigma^2) & \text{with probability } \rho \\ N \left(0, \frac{1}{\lambda} \sigma^2 \right) & \text{with probability } 1 - \rho \end{cases},$$

where $0 < \lambda < 1$, and $\sigma^2 = (\rho + (1 - \rho)/\lambda)^{-1}$ so that $\text{var}(\varepsilon_t) = 1$. The innovations $\varepsilon_t$ are generated from a normal distribution with variance $\sigma^2$ with probability $\rho$, or from a normal distribution with variance $\sigma^2/\lambda$ with probability $1 - \rho$. This model allows for conditional nonnormality and explains the leptokurtosis observed in financial series.
The kurtosis of the innovations is given by

\[ K_e = \frac{3(1 - \rho)(1/(\hat{\lambda}) - 1)^2}{(\rho + (1 - \rho)(1/\hat{\lambda}))^2}. \]

The proposed TNM-GARCH model is given by

\[ y_t = m_t + n_t, \quad t = 1, \ldots, N, \]

where the conditional mean specification \( m_t \) allows for different AR structures in each partition of the threshold model. Usually, low-order AR structures are able to capture the weak dependencies appeared in financial series. The conditional mean is given by

\[ m_t = \sum_{j=1}^{k} \left( \mu_j + \sum_{i=1}^{r} \phi_{ji} y_{t-i} \right) I(y_{t-1} \in P_j), \]

while \( n_t \) is given by

\[ n_t = \sqrt{h_t} \sum_{j=1}^{k} \varepsilon_{jt} I(y_{t-1} \in P_j), \]

with

\[ \varepsilon_{jt} \sim \begin{cases} N(0, \sigma_j^2) & \text{with probability } \rho_j \\ N \left( 0, \frac{1}{\lambda_j} \sigma_j^2 \right) & \text{with probability } 1 - \rho_j \end{cases}, \]

and \( 0 < \lambda_j < 1, \sigma_j^2 = (\rho_j + (1 - \rho_j)/\lambda_j)^{-1} \) so that \( \text{var}(\varepsilon_{jt}) = 1 \). Thus the innovations \( \varepsilon_{jt} \) in the \( j \)th partition \( P_j \) are generated from a normal distribution with variance \( \sigma_j^2 \) with probability \( \rho_j \), or from a normal distribution with variance \( \sigma_j^2/\lambda_j \) with probability \( 1 - \rho_j \). \( h_t \) is the conditional variance of \( y_t \) given the information set up to time \( t-1 \), and is given by a GARCH specification

\[ h_t = \omega + \sum_{i=1}^{p} a_i n_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \]

where \( \omega > 0, a_i \geq 0, i = 1, \ldots, p, \beta_j \geq 0, j = 1, \ldots, q \) to ensure positivity of \( h_t \). To identify the model and keep the number of parameters relatively small, the parameters in the variance equation are not state (partition) dependent. The parameter vector to be estimated in the NM-GARCH specification is \( \theta_{\text{NM}} = (\theta_{\text{NM},1}, \ldots, \theta_{\text{NM},k}, \theta_{\text{G}}) \), where \( \theta_{\text{NM},j} = (\mu_j, \phi_{j1}, \ldots, \phi_{jr}, \lambda_j, \rho_j), j = 1, \ldots, k \) and \( \theta_{\text{G}} = (\omega, a_1, \ldots, a_p, \beta_1, \ldots, \beta_q) \). The parameter vector \( \theta \) of the proposed TNM-GARCH model includes the vector of parameters \( (k, \theta_T^{(k)}) \) of the threshold structure and the vector of the NM-GARCH parameters \( \theta_{\text{NM}}, \) so \( \theta = (k, \theta_T^{(k)}, \theta_{\text{NM}}) \). Inference for the parameter vector \( \theta \) is achieved conditional on initial observations that are used to determine the partitions \( P_j \).

The proposed TNM-GARCH model can be considered as a generalization of the NM model introduced by Bai et al. (2003) and further used by Ausin and Galeano (2007) to successfully capture the leptokurtosis observed in financial time series. It is obvious that if there is no threshold in the model (i.e. \( k = 1 \)) it reduces to the pure NM model of Bai et al. (2003). On the other hand, if there is no mixture in the model (i.e. under conditional normality) it reduces to a threshold in mean model with a GARCH-type conditional variance. The proposed model specification can capture several stylized facts such as volatility clustering, high kurtosis, heavy tails and the presence of extreme events. The threshold structure allows modelling the nonlinear behaviour of financial series, since it seems natural to allow for the existence of different states or partitions, and to model the underlying dynamics in a different way. In particular, it allows the conditional mean and kurtosis (fourth moment) to vary (change) between different partitions \( P_j \) since the parameters \( \theta_{\text{NM},j} \) are different.
2.2.2. The threshold mixed normal conditional heteroscedasticity model

The general NM model of Hass et al. (2004) can be written in the form

\[ y_t = m_t + e_t, \quad t = 1, \ldots, N, \]

where \( m_t \) specifies the conditional mean of the process, and \( e_t \) is the error term. Hass et al. (2004) used common mean dynamics for the conditional mean equation based on ARMA structure, and an \( s \)-component mixed normal distribution for the error process together with a GARCH-type structure which allows for conditional variances in each of the components. That is

\[ e_t | \Phi_{t-1} \sim \text{MN}(p_1, \ldots, p_s, \mu_1, \ldots, \mu_s, h_{1t}, \ldots, h_{st}), \]

where \( \Phi_{t-1} \) is the information set up to time \( t - 1 \), \( p_i \in (0, 1) \), \( i = 1, \ldots, s \), \( \sum_{i=1}^{s} p_i = 1 \) and \( \mu_s = -\sum_{i=1}^{s-1} (p_i / p_s) \mu_i \). Moreover, the vector of component variances is given by

\[ h_t = \omega + \sum_{i=1}^{p} a_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \]

where \( h_t = (h_{1t}, \ldots, h_{st})' \), \( \omega = (\omega_1, \ldots, \omega_s)' \), \( a_i = (a_{i1}, \ldots, a_{is})' \), \( i = 1, \ldots, p \) and \( \beta_j, j = 1, \ldots, q \) are \( s \times s \) matrices. A diagonal mixture normal specification has diagonal \( \beta_j \) matrices.

In this section we present a threshold mixed normal conditional heteroscedasticity model (TMNCH hereafter) that generalizes the model of Haas et al. (2004) by allowing the conditional variance dynamics to change in each partition. That is the proposed model can be written as

\[ y_t = m_t + n_t, \quad t = 1, \ldots, N, \]

where common mean dynamics are used in the conditional mean equation and \( n_t \) is given by

\[ n_t = \sum_{j=1}^{k} e_{jt} I(y_{t-1} \in P_j). \]

The conditional distribution of \( e_{jt} \), in partition \( P_j \), is an \( s \)-component normal mixture with mean zero, i.e.

\[ e_{jt} | \Phi_{t-1} \sim \text{MN}(p_{j1}, \ldots, p_{js}, \mu_{j1}, \ldots, \mu_{js}, h_{j1t}, \ldots, h_{jst}), \]

where \( \Phi_{t-1} \) is the information set up to time \( t - 1 \), \( p_{ji} \in (0, 1) \), \( i = 1, \ldots, s \), \( j = 1, \ldots, k \), \( \sum_{i=1}^{s} p_{ji} = 1 \) and \( \mu_{js} = -\sum_{i=1}^{s-1} (p_{ji} / p_{js}) \mu_{ji} \). Moreover, the conditional variances using a diagonal MN specification are given by

\[ h_{jlt} = \omega_{ji} + \sum_{l=1}^{p} a_{jli} n_{t-l}^2 + \sum_{m=1}^{q} \beta_{jim} h_{j,t-m}, \quad i = 1, \ldots, s, \quad j = 1, \ldots, k, \]

where \( \omega_{ji} > 0, a_{jli} \geq 0, \beta_{jim} > 0, i = 1, \ldots, s, \quad j = 1, \ldots, k, \quad l = 1, \ldots, p, \quad m = 1, \ldots, q \). The parameter vector to be estimated in the MNCH specification is \( \theta_{\text{MNCH}} = (\theta_{\text{MNCH},1}, \ldots, \theta_{\text{MNCH},k}, \theta_{\text{AR}}) \), where \( \theta_{\text{AR}} = (\mu, \phi_1, \ldots, \phi_p) \) is the vector of parameters in the common conditional mean specification and \( \theta_{\text{MNCH},j} = (p_{j1}, \ldots, p_{js}, \mu_{j1}, \ldots, \mu_{j,s-1}, \omega_{j1}, \ldots, \omega_{js}, a_{j1l}, \ldots, a_{jsl}, \beta_{j1m}, \ldots, \beta_{jsm}) \), \( j = 1, \ldots, k \). The parameter vector \( \theta \) of the proposed TMNCH-GARCH model includes the vector of parameters \( k \) of the threshold structure and the vector \( \theta_{\text{MNCH}} \) of the MNCH specification, thus \( \theta = (k, \theta^T, \theta_{\text{MNCH}}) \). The above specification has a large number of parameters and is more complicated and computationally demanding than the TMN model. Inference for the parameter vector \( \theta \) is achieved conditionally on initial observations that are used to determine the partitions \( P_j \). Note that if there is no threshold in the model \( (k = 1) \) the TMNCH-GARCH model reduces to the MNCH model of Haas et al. (2004). It is obvious that the proposed model specification can capture several nonlinearities of financial series, since it allows the conditional variances to vary (change) between different partitions \( P_j \).
2.3. Specifying prior probabilities

The proposed model is identified by a parameter vector $\theta = (k, \theta_T^{(k)}, \theta_m)$, where $\theta_m$ denotes the parameter vector of the mixture normal specification, i.e. either of the NM-GARCH model or of the MNCH model. Before implementing the Bayesian methodology, the prior distribution $p(k, \theta_T^{(k)}, \theta_m)$ on the parameters must be specified. It is convenient to use

$$p(k, \theta_T^{(k)}, \theta_m) = p(\theta_m | k, \theta_T^{(k)}) p(k, \theta_T^{(k)})$$

and specify $p(k, \theta_T^{(k)})$ and $p(\theta_m | k, \theta_T^{(k)})$ separately. The advantage of this specification is that the choice of prior for the model $T$ with $k$ partitions does not depend on the form of the parametric family indexed by $\theta_m$. The conditional specification of the prior on $\theta_m$ also facilitates the posterior computation.

For the prior specification of the threshold-model parameters we use a hierarchical structure

$$p(k, \theta_T^{(k)}) = p(k)p(\theta_T^{(k)} | k) = p(k)p(t_1, t_2, \ldots, t_{k-1} | k).$$

For the prior density of the number of partitions $k$, we use a Poisson distribution truncated to one with p.d.f.

$$p(k) = \frac{k^k}{(e^\gamma - 1)k!}, \quad \gamma > 0, \quad k = 1, 2, \ldots$$

so that we require a minimum of one partition in the model. Then, conditional on $k$, we assign equal prior probabilities to all models with $k$ partitions. That is, if $n$ denotes the number of potential thresholds, we use a discrete uniform prior distribution over the set of $(\frac{n}{k-1})$ possible ways of picking $k - 1$ out of the $n$ thresholds, that is, $p(t_1, t_2, \ldots, t_{k-1} | k) = [(k - 1)!/(n - (k - 1))]/(n!)$.

For the parameters of the NM-GARCH model, $\theta_{NM}$, we assume proper prior distributions to avoid resulting in an improper posterior. We consider two different prior specifications. The first assumes a noninformative prior distribution with large prior variance. In order to avoid imposing positivity restrictions to the parameters in the conditional variance and constraints to the parameters $\lambda_j$ and $\rho_j$, we transform those parameters to a subvector taking values in $(-\infty, \infty)$. In particular, we use the logarithmic transformation, $\omega^* = \ln(\omega), a_i^* = \ln(a_i), i = 1, \ldots, p$ and $b_m^* = \ln(b_m)$, $j = 1, \ldots, q$, for the positive parameters and the logit transformation $\rho_j^* = \ln(\rho_j/(1 - \rho_j))$ and $\lambda_j^* = \ln(\lambda_j/(1 - \lambda_j))$ for the parameters $\rho_j$ and $\lambda_j$, $j = 1, \ldots, k$. For the autoregressive parameters we use the reparameterization strategy discussed in Marriott et al. (1996). For example, in the case of the AR(1) model we use the transformation $\phi_j^* = \ln((1 + \phi_j)/(1 - \phi_j))$. It appears that it is computationally more convenient to work with the transformed parameters; see for example, Miazhynskaia and Dorfner (2006), and Vrontos et al. (2003). We adopt proper normal prior distributions for the transformed parameters of the form

$$\rho_j^* \sim N(\mu_{\rho_j^*}, \sigma_{\rho_j^*}^2), \quad \lambda_j^* \sim N(\mu_{\lambda_j^*}, \sigma_{\lambda_j^*}^2), \quad \mu_j \sim N(\mu_{\mu_j}, \sigma_{\mu_j}^2), \quad \phi_j^* \sim N(\mu_{\phi_j^*}, \sigma_{\phi_j^*}^2),$$

$$\omega^* \sim N(\mu_{\omega^*}, \sigma_{\omega^*}^2), \quad a_i^* \sim N(\mu_{a_i^*}, \sigma_{a_i^*}^2), \quad b_m^* \sim N(\mu_{b_m^*}, \sigma_{b_m^*}^2),$$

$j = 1, \ldots, k, i = 1, \ldots, r, l = 1, \ldots, p$ and $m = 1, \ldots, q$, where the parameters of the prior distributions are chosen to reflect prior expectation/evidence from the empirical application of this type of models. In both the simulation and real data examples these hyperparameters were set as $\mu_{\rho_j^*} = 2.5, \mu_{\lambda_j^*} = -1.73, \mu_{\mu_j} = 0, \mu_{\phi_j^*} = 0, \mu_{\omega^*} = -11, \mu_{a_i^*} = -2.5, \mu_{b_m^*} = -0.1, \sigma_{\rho_j^*} = \sigma_{\lambda_j^*} = 2, \sigma_{\mu_j} = \sigma_{\phi_j^*} = 1, \sigma_{\omega^*} = 3, \sigma_{a_i^*} = 2$ and $\sigma_{b_m^*} = 1$. The dispersion of these prior distributions turned out to be much larger than that of the corresponding posterior distributions. The second prior distribution is based on the unit information prior proposed by Kass and Wasserman (1995) and further applied by, for example, Dellaportas and Forster (1999) and Dellaportas and Vrontos (2005). We use a multivariate normal distribution $N(\mu_j, N_j\tilde{\Sigma}_j)$ for the parameters $\theta_{NM, j} = (\mu_j, \phi_j^*, \lambda_j^*, \rho_j^*)$ of the $j$th partition, where $\mu_j = (\mu_{\mu_j}, \mu_{\phi_j^*}, \mu_{\lambda_j^*}, \mu_{\rho_j^*})$ is the prior mean, and the covariance matrix $N_j\tilde{\Sigma}_j$ depends on the number of data points $N_j$ that belong to the $j$th partition and on the covariance matrix $\tilde{\Sigma}_j$ of the maximum likelihood estimates of this parameters. For the parameters in the conditional variance $\theta_{G}^* = (\omega^*, a_i^*, b_m^*)$ we adopt a unit information prior distribution by using a multivariate normal distribution with mean $(\mu_{\omega^*}, \mu_{a_i^*}, \mu_{b_m^*})$ and covariance matrix that depends on the covariance matrix of the maximum likelihood
estimates and on the number of data points used in the analysis. For the parameters of the MNCH-GARCH model, \( \theta_{\text{MNCH}} \), we consider similar prior specifications, i.e. a noninformative prior distribution with large prior variance and a unit information prior.

3. Bayesian inference under model uncertainty

3.1. Posterior model probabilities

As described in Section 2, the number of partitions \( k \) and the parameter vector \( \theta_{T}^{(k)} \) of a model \( T \) with \( k \) partitions determine uniquely the threshold structure, which can be considered as a specific model denoted by \( T_{k, \theta} \). Inference on the parameter vector \( \theta_{m} \) of a given model \( T_{k, \theta} \) can be easily made using either a classical or a Bayesian approach. The maximum likelihood estimate of \( \theta_{m} \) can be obtained by maximizing the log-likelihood of the observed data using some numerical maximization technique. We use the quasi-Newton maximization method available in Matlab. Within the Bayesian framework all inferences about the model parameters \( \theta_{m} \) are summarized by the posterior distribution \( p(\theta_{m}|y, T_{k, \theta}) \). To sample draws from this posterior distribution we follow the approach of Vrontos et al. (2003) who use a multivariate random walk Metropolis–Hastings algorithm which simultaneously updates the components of the parameter vector \( \theta_{m} \). That is, at the \( i \)th iteration of the MCMC algorithm, we propose a candidate parameter vector \( \theta_{m}^{\text{can}} \) from a multivariate normal density \( N(\theta_{m}, \Sigma) \), where \( \theta_{m} \) denotes the current parameter value, \( \Sigma \) is an estimate of the parameter covariance matrix based on maximum likelihood and \( c \) is a constant which tunes the acceptance rate.

Since there are a number of different competing models, corresponding to different threshold structures, it is of great interest to perform some model selection exercise. A typical approach of greedy algorithms is to carry out model selection based on some penalty function, leading to a single ‘best’ threshold structure, and then make inference as if the selected model was the true model. However, this ignores the uncertainty involved in the model selection exercise. A Bayesian solution to this problem involves the calculation of the posterior probabilities of all competing threshold models. The posterior probability of model \( T_{k, \theta} \) is given by

\[
p(T_{k, \theta}|y) = \frac{p(y|T_{k, \theta})p(T_{k, \theta})}{\sum p(y|T_{k, \theta})p(T_{k, \theta})}, \tag{1}
\]

where

\[
p(y|T_{k, \theta}) = \int p(y|T_{k, \theta}, \theta_{m})p(\theta_{m}|T_{k, \theta}) \, d\theta_{m}
= \int p(y|k, \theta_{T}^{(k)}, \theta_{m})p(\theta_{m}|k, \theta_{T}^{(k)}) \, d\theta_{m} \tag{2}
\]

is the marginal likelihood for model \( T_{k, \theta} \) with \( k \) partitions and threshold parameters \( \theta_{T}^{(k)}, \theta_{m} \) is the parameter vector of the normal mixture specifications in threshold model \( T_{k, \theta} \). \( p(y|T_{k, \theta}, \theta_{m}) \) or \( p(y|k, \theta_{T}^{(k)}, \theta_{m}) \) is the likelihood given the threshold model \( T_{k, \theta} \). \( p(\theta_{m}|T_{k, \theta}) \) or \( p(\theta_{m}|k, \theta_{T}^{(k)}) \) is the conditional specification of the prior on \( \theta_{m} \) under threshold model \( T_{k, \theta} \), and \( p(T_{k, \theta}) \) or \( p(k, \theta_{T}^{(k)}) \) is the prior probability for threshold model \( T_{k, \theta} \).

The marginal likelihood \( p(y|T_{k, \theta}) \) can be estimated using a variant of Laplace’s method of approximation given by

\[
p(\hat{y}|T_{k, \theta})_{\text{mle}} = (2\pi)^{d_{\theta_{m}}/2} |\hat{\Sigma}|^{1/2} p(y|\hat{\theta}_{m}, T_{k, \theta}) p(\hat{\theta}_{m}|T_{k, \theta}), \tag{3}
\]

where \( \hat{\Sigma} \) is the inverse of the negative Hessian matrix of the log-likelihood evaluated at the maximum likelihood estimator \( \hat{\theta}_{m} \) and \( p(\hat{\theta}_{m}|T_{k, \theta}) \) are the likelihood and the prior, respectively, evaluated at \( \hat{\theta}_{m} \). This method of approximation of the marginal likelihood will be used in the stochastic search algorithm described below; it is sufficiently accurate to be of use in many problems (Kass and Raftery, 1995) and has been adopted by Vrontos et al. (2003) and Dellaportas and Vrontos (2005) in the analysis of multivariate time varying volatility models. We use this method of estimating the marginal likelihood within our stochastic search algorithm described in the next section.
3.2. MCMC stochastic search algorithm

Analytic evaluation of (1) over all threshold models \( T_{k, \theta} \) is computationally infeasible, because of the large number of possible thresholds. We describe an MCMC algorithm that jumps between models of the same or different dimensionality to generate a sample from the posterior distribution \( p(k, \theta^{(k)}_T) \), or in other words to provide posterior probabilities for the threshold models \( T_{k, \theta} \).

Metropolis–Hastings-type algorithms can be used to explore the posterior distribution. These algorithms simulate a Markov chain consisting of threshold models, \( T^{(0)}_{k, \theta}, T^{(1)}_{k, \theta}, T^{(2)}_{k, \theta}, \ldots \), which, under weak conditions (see Tierney, 1994), converges to the limiting distribution \( p(T_{k, \theta}|y) \). The Metropolis–Hastings algorithm for simulating \( T^{(0)}_{k, \theta}, T^{(1)}_{k, \theta}, T^{(2)}_{k, \theta}, \ldots \) is defined as follows. Starting with an initial model \( T^{(0)}_{k, \theta} \), iteratively simulate the transitions from \( T^{(i)}_{k, \theta} \) to \( T^{(i+1)}_{k, \theta} \) by using two steps:

- **Generate a candidate model** \( T^{(i+1)}_{k', \theta} \) with probability distribution \( q(T^{(i)}_{k, \theta}, T^{(i+1)}_{k', \theta}) \).
- **Accept** \( T^{(i+1)}_{k, \theta} \) with probability

\[
\alpha(T^{(i)}_{k, \theta}, T^{(i+1)}_{k', \theta}) = \min \left\{ \frac{q(T^{(i)}_{k', \theta}, T^{(i)}_{k, \theta}) \cdot p(y|T^{(i)}_{k', \theta}) \cdot p(T^{(i)}_{k', \theta})}{q(T^{(i)}_{k, \theta}, T^{(i+1)}_{k', \theta}) \cdot p(y|T^{(i)}_{k, \theta}) \cdot p(T^{(i+1)}_{k', \theta})}, 1 \right\}, \tag{4}
\]

otherwise set \( T^{(i+1)}_{k, \theta} = T^{(i)}_{k, \theta} \). Note that the normalizing constant for \( p(T_{k, \theta}|y) \) is not needed to compute (4).

In our study, since we deal with models \( T_{k, \theta} \) which have parameter vector \((k, \theta^{(k)}_T)\) of different dimensionality, we shall need to devise different types of moves between the subspaces. Therefore, we consider transition kernels \( q(T^{(i)}_{k, \theta}, T^{(i+1)}_{k', \theta}) \) which generate candidate threshold models \( T^{(i+1)}_{k', \theta} \) from current models \( T^{(i)}_{k, \theta} \) by randomly choosing among the following steps:

- **Birth**: Randomly choose a new threshold (generates a new partition).
- **Death**: Randomly pick a threshold in the model and delete it (kills an existing partition).
- **Change-threshold**: Randomly pick a threshold and replace with a new threshold from the available ones.

Note that the **Birth** and **Death** steps are the reverse of each other and both of them change the dimension of the model parameter vector, while the **Change-threshold** step proposes jumps within a given dimension. To simulate from the posterior density of thresholds, **Birth** and **Death** moves are required since they change the number of partitions \( k \) (increase or decrease \( k \) by one) and the corresponding parameter vector \( \theta^{(k)}_T \). The **Change-threshold** move proposes local jumps by changing the threshold value of the model.

3.3. Implementation details

To clarify the proposed algorithm, we describe the computation of the acceptance probability \( \alpha \) in (4) for the moves that change the threshold-model dimensionality. Consider a birth move from a current threshold-model \( T^{(i)}_{k, \theta} \) with \( k \) partitions and threshold parameters \( \theta^{(k)}_T \), to a proposed new threshold-model \( T^{(i+1)}_{k', \theta} \) with \( k'=k+1 \) partitions and threshold parameters \( \theta^{(k')}_T \). Suppose \( n \) is the number of potential thresholds, so that the model will have at most \( k_{\text{max}} = n + 1 \) partitions. The probability of acceptance (4) can be written as

\[
\alpha = \min\{\text{marginal likelihood ratio} \times \text{proposal ratio} \times \text{prior ratio}, 1\},
\]

where the marginal likelihood can be calculated using the Laplace method, and the prior of a threshold-model \( T \) with \( k \) partitions, \( p(T_{k, \theta}) \), is given by \( p(k) p(\theta^{(k)}_T | k) \).
Let $b_k$ be the probability of proposing a birth move, $d_k$ be the probability of proposing a death move, and $c h_k$ the probability of changing a threshold. The probability of moving from the proposed threshold-model $T_k(i)$ to the current threshold-model $T_k$, is $q(T_k(i), T_k)$ and is given by
\[
q(T_k(i), T_k) = d_k + 1 \frac{1}{k},
\]
i.e. it is the probability of performing a death step, $d_{k+1}$, multiplied by the probability of choosing a threshold to delete in order to form a threshold-model with one partition less. Similarly, the probability of moving from the current threshold-model $T_k(i)$ to the proposed threshold-model $T_k(i)'$, with one more partition (one more threshold) is given by
\[
q(T_k(i), T_k(i)') = b_k p(t_k | t_1, \ldots, t_{k-1}, k + 1) = b_k (n - (k - 1))^{-1},
\]
which is the product of the probability of performing a birth step, $b_k$, times the probability of proposing a specific threshold from those remaining. Therefore, for the birth move, the proposal multiplied by the prior ratio equals
\[
\frac{p(k + 1) d_{k+1}}{p(k) b_k}.
\]
For a death step the above ratio is inverted. For the steps which involve a change in the threshold structure the acceptance probability is easier to calculate since the proposal ratio $\times$ prior ratio cancel each other.

Algorithm.

1. Begin with a model without thresholds (i.e. $k = 1$).
2. Generate $u$ uniformly on $[0, 1]$.
3. The move type is determined by $u$
   - if ($u \leq b_k$) then propose a Birth step
   - if ($b_k < u \leq b_k + d_k$) then propose a Death step
   - else propose a Change step.
4. Calculate the probability of acceptance, and set $k$ equal to the number of partitions in the present threshold-model.
5. Repeat 2–4 until convergence, for example until only very little change in the posterior probability of the threshold-model structure is observed.

3.4. Model averaging

Being able to calculate the posterior probabilities (1) of each threshold model $T_k$, it seems natural to account for model uncertainty in our predictive inferences. Rather than choosing a single ‘best’ model and then making inferences as if the selected model was the true one, we can use the following model averaging approach, which provides composite predictions. Suppose that we are interested in a quantity $A$. For example, in financial models this quantity may be a future observation or the variance at a future time period. Then, its posterior distribution given data $y$ is given by
\[
p(A|y) = \sum_{T_k \in A} p(A|T_k, y) p(T_k|y),
\]
which is an average of the posterior predictive distribution under each model weighted by their posterior probabilities, and $A$ is the set of all $2^{k_{\text{max}} - 1}$ possible models. The posterior predictive distribution of $A$ given a particular model $T_k$ is found by integrating out the model parameters $\theta_m$ of the specific model
\[
p(A|T_k, y) = \int p(A|\theta_m, T_k, y) p(\theta_m|T_k, y) d\theta_m.
\]
We can also use the maximum likelihood approximation

\[
p(A|T_k, \theta, y) \approx p(A|T_k, \theta, \hat{\theta}_m),
\]

where \( \hat{\theta}_m \) is the maximum likelihood estimator of the parameter vector \( \theta_m \) of threshold model \( T_k, \theta \). For a discussion of the above approach as well as for evidence on that accounting for model uncertainty improves predictive performance see Kass and Raftery (1995) and Raftery et al. (1997).

4. Simulation study

In this section, we conduct a simulation study to investigate the performance of the Bayesian approach suggested in the previous sections to detect the correct data generating process, that is to identify the specific threshold model that was used to simulate the data. We have implemented an extensive simulation study considering different scenarios, using different time series lengths and various threshold structures.

Some of the results of that simulation study are summarized in Table 1. First, we have considered the TNM-GARCH model (Panel A) with a conditional mean specification that allows for different AR(1) structures in each partition of the model and a GARCH(1,1) specification for the variance equation. We applied the MCMC stochastic search algorithm for 100,000 iterations using the prior distributions described in Section 2.3. We used the empirical percentiles (10th, 20th, . . . , 99th) of the simulated data as possible threshold values, i.e. \( k_{\text{max}} = 10 \). To investigate the ability of our algorithm to identify the correct threshold values \( (t_1, t_2, \ldots, t_k) \) we replaced some of the percentiles by their respective nearby true threshold values used in the simulation. The method seems to be flexible and efficient since it arrives at the correct threshold structure very fast. Note also that the MCMC stochastic search algorithm has very good convergence properties. In all simulations, the estimated posterior probabilities of the true threshold models were greater than 0.99 (see Table 1, column 4).

For example, in the fifth simulation scenario of Table 1, we generated \( N = 5000 \) data points from a TNM-GARCH model with three threshold values \( t_1 = -0.03, t_2 = 0 \) and \( t_3 = 0.04 \), which subdivide the observation space into \( k = 4 \) partitions. The stochastic search algorithm identified the true threshold structure, i.e. the correct threshold values \( t_1, t_2 \) and \( t_3 \) from a set of possible thresholds, as well as the number of partitions, with posterior probability 0.999. Running this algorithm for 100,000 iterations took about 260 min using Matlab on a Pentium D processor.

In Fig. 1(a) and (b) we present the simulated data and the corresponding simulated conditional variances, while Fig. 1(c) illustrates the estimated posterior mean of the conditional variances based on the MCMC sample indicating the accuracy of the Bayesian variance estimates. In Table 2 we present the parameter values that were used in this simulation scenario together with the estimates and standard deviations/errors using Bayesian and classical techniques. Fig. 2 shows the histograms and the posterior densities of the model parameters based on the MCMC output. It can be seen that the Bayesian posterior means and the classical point estimates are very similar, which can be explained by

---

Table 1

<table>
<thead>
<tr>
<th>N</th>
<th>k</th>
<th>Threshold values: ( \theta_f^{(k)} )</th>
<th>PP</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: TNM-GARCH specification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>( t_1 = -0.007, t_2 = 0.037 )</td>
<td>0.998</td>
<td>27</td>
</tr>
<tr>
<td>1000</td>
<td>4</td>
<td>( t_1 = -0.04, t_2 = 0, t_3 = 0.05 )</td>
<td>0.991</td>
<td>54</td>
</tr>
<tr>
<td>3000</td>
<td>3</td>
<td>( t_1 = -0.03, t_2 = 0.03 )</td>
<td>0.999</td>
<td>55</td>
</tr>
<tr>
<td>3000</td>
<td>4</td>
<td>( t_1 = -0.04, t_2 = 0, t_3 = 0.05 )</td>
<td>0.997</td>
<td>134</td>
</tr>
<tr>
<td>5000</td>
<td>4</td>
<td>( t_1 = -0.03, t_2 = 0, t_3 = 0.04 )</td>
<td>0.999</td>
<td>260</td>
</tr>
<tr>
<td><strong>Panel B: TMNCH-GARCH specification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>2</td>
<td>( t_1 = 0.3 )</td>
<td>0.999</td>
<td>602</td>
</tr>
<tr>
<td>5000</td>
<td>2</td>
<td>( t_1 = 0.3 )</td>
<td>0.999</td>
<td>1026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>k</th>
<th>Threshold values: ( \theta_f^{(k)} )</th>
<th>PP</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| N is the number of simulated observations, k is the number of partitions, \( \theta_f^{(k)} \) is the vector with the threshold values, PP denotes the posterior probability of the true model-threshold structure and CT denotes the computing time (in minutes) for 100,000 iterations.
the fact that the posterior distributions of the transformed parameters are bell-shaped (see Fig. 2). The advantage of the Bayesian approach is that it provides posterior densities for the model parameters, which capture all the uncertainty associated with them, rather than producing simple point estimates and standard errors based on asymptotic results.

To examine the robustness of the results to the choice of prior distributions we have conducted a sensitivity analysis; we used the unit information prior for the model parameters, and we considered two different values for the parameter of the truncated Poisson distribution on the number of partitions $k$, namely $\gamma = 2$ and 3. The stochastic search algorithm successfully identified the true generating process with posterior probabilities above 0.99 using all different prior specifications.

We have also considered simulation scenarios using a two component TMNCH-GARCH model with a common AR(1) structure for the conditional mean and a different GARCH(1,1) specification for the conditional variances in each component. Table 1 (panel B) presents results for two such simulation scenarios. The proposed algorithm produces posterior probabilities of the true models above 0.99 for both scenarios, using all different prior specifications. We observe from Table 1, that the computational time required to run the algorithm for 100,000 iterations for this simulation scenario is larger than that of the TNM-GARCH model. Obviously, the computational time increases with the length of the time series. The parameters used in the first simulation scenario, the estimates and the standard errors are illustrated in Table 3.

Note that the simple MNCH-GARCH model has a more complicated structure than the simple NM model and, therefore, incorporating the former within the threshold setting produces a model with a quite heavy structure. The complicated form of this model increases the computational complexity of our algorithm; the computational cost increases substantially and occasionally numerical problems due to overfitting may occur (especially if the model imposes too much structure to the data at hand).
Table 2
Simulated and estimated parameter values of the TNM-GARCH model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True values</th>
<th>Transformed</th>
<th>Bayesian estimates</th>
<th>Classical estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_1 )</td>
<td>0.78</td>
<td>( \rho_1^* )</td>
<td>1.27</td>
<td>1.18</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>0.06</td>
<td>( \lambda_1^* )</td>
<td>-2.75</td>
<td>-2.70</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>-0.04</td>
<td>( \mu_1^* )</td>
<td>-0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>0.1</td>
<td>( \phi_1^* )</td>
<td>0.20</td>
<td>0.15</td>
</tr>
<tr>
<td>( \rho_2 )</td>
<td>0.805</td>
<td>( \rho_2^* )</td>
<td>1.42</td>
<td>1.66</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>0.08</td>
<td>( \lambda_2^* )</td>
<td>-2.44</td>
<td>-2.60</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>-0.015</td>
<td>( \mu_2^* )</td>
<td>-0.015</td>
<td>-0.017</td>
</tr>
<tr>
<td>( \phi_2 )</td>
<td>0.3</td>
<td>( \phi_2^* )</td>
<td>0.62</td>
<td>0.49</td>
</tr>
<tr>
<td>( \rho_3 )</td>
<td>0.83</td>
<td>( \rho_3^* )</td>
<td>1.59</td>
<td>1.53</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>0.1</td>
<td>( \lambda_3^* )</td>
<td>-2.20</td>
<td>-2.02</td>
</tr>
<tr>
<td>( \mu_3 )</td>
<td>0.01</td>
<td>( \mu_3^* )</td>
<td>0.01</td>
<td>0.008</td>
</tr>
<tr>
<td>( \phi_3 )</td>
<td>0.2</td>
<td>( \phi_3^* )</td>
<td>0.41</td>
<td>0.76</td>
</tr>
<tr>
<td>( \rho_4 )</td>
<td>0.855</td>
<td>( \rho_4^* )</td>
<td>1.77</td>
<td>1.80</td>
</tr>
<tr>
<td>( \lambda_4 )</td>
<td>0.12</td>
<td>( \lambda_4^* )</td>
<td>-1.99</td>
<td>-1.91</td>
</tr>
<tr>
<td>( \mu_4 )</td>
<td>0.05</td>
<td>( \mu_4^* )</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( \phi_4 )</td>
<td>0.3</td>
<td>( \phi_4^* )</td>
<td>0.62</td>
<td>0.57</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.0002</td>
<td>( \omega^* )</td>
<td>-8.52</td>
<td>-8.71</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.1</td>
<td>( \alpha_1^* )</td>
<td>-2.30</td>
<td>-2.25</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.85</td>
<td>( \beta_1^* )</td>
<td>-0.16</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

N = 5000 observations, \( k = 4 \) partitions, \( \theta_1^{(k)} = (t_1 = -0.03, t_2 = 0, t_3 = 0.04) \). PM denotes the posterior mean, PS denotes the posterior standard deviation, Est denotes estimates and StE denotes standard errors.

5. Application of the proposed model

We apply the TNM-GARCH models to analyse daily observations on the S&P 500 composite price index over the 14/7/1989–1/4/2005 period. Let \( S_t \) denote the value of the index at time \( t \). Then we model the rates of return \( y_t = \ln(S_t / S_{t-1}) \), \( t = 1, \ldots, N = 4100 \). The total sample of 4100 rates of returns is split into two parts: the first 3300 data points are used to make inference about the threshold structures and to estimate their posterior probabilities (estimation period). The last 800 data points can be used to examine the predictive performance of the threshold structure (out of sample performance period). Our application is concerned with the calculation of the Value at Risk (VaR) that is widely used in the evaluation of risk management. GARCH-type models have been used in the literature to provide accurate VaR forecast; see, for example, the paper of Hartz et al. (2006). In Table 4, we present the summary statistics for the rates of return of the analysed index (for the estimation period) together with the Ljung–Box statistics computed for the rates of return \( y_t \), for the absolute rates, and for the squares of the rates. The Ljung–Box statistic is computed using 50 lags and shows autocorrelation in the returns and high level of autocorrelation in the squares values of rates of return and mainly in the absolute values. The kurtosis of the return series is 7.8157, indicating fat-tailness in the returns distribution. In Fig. 3(a), we present the analysed rates of return for the S&P 500 index. The volatility clustering phenomenon is clear in the return series.

5.1. Inference for the threshold structure

In this section we present results for S&P 500 index obtained from the application of our MCMC stochastic search algorithm to two model specifications; the TNM-GARCH model and the TMNCH-GARCH model. In both cases, we have used an AR model of order one, AR(1), in the conditional mean equation and a GARCH(1,1) specification to model the conditional variances. This structure can usually capture the autocorrelation in returns and squared returns of financial series. Our approach enables us to identify the most probable models (in each specification), and calculate their posterior model probabilities. The marginal likelihoods of these models, needed for the calculation of the posterior
Fig. 2. Histograms and posterior densities of the MCMC output for each parameter.

probabilities, can be further used to compare the most probable models under the two different specifications. Then, based on a single ‘best’ model, we can make inferences about the model parameters and other quantities of interest such as future variances. Moreover, robust inferences about such quantities can be drawn within the Bayesian framework by using BMA, a technique which accounts for model uncertainty or miss-specification.

First, we applied the proposed Bayesian approach to the TNM-GARCH specification. We run the MCMC stochastic search algorithm for 100,000 iterations, using the empirical percentiles of the observed series (10th, 20th, ...) as the set of possible threshold values i.e. $k_{\text{max}} = 10$ (see Table 5). We considered two different prior specifications, i.e. the independent noninformative prior and the unit information prior using the same hyperparameter values as in the simulation study. We also used two different values for the parameter of the truncated Poisson prior, i.e. $\gamma = 2$ and $3$ to examine the sensitivity of the results to the prior information for the number of partitions $k$. In Fig. 4 are plotted the posterior probabilities of the different threshold models that the algorithm has visited across the sweeps calculated ergodically every 1000 iterations, obtained using the four different prior specifications. Note that this is a conservative run; the behaviour of the MCMC stochastic search chain was extremely good with rapid convergence and therefore only a half or one-third of the iterations would be adequate. The method seems to be very efficient since the algorithm arrives at the ‘best’ model very fast. The posterior probabilities seem to be robust to different prior specifications.

Next we applied our stochastic search algorithm to the two component TMNCH-GARCH model. Again we used two different prior specifications, i.e. the noninformative and the unit information prior, for MNCH-GARCH parameters and the values $\gamma = 2$ and $3$ for the parameter of the truncated Poisson prior. Table 6 presents the six most probable models within the TNM-GARCH specification and the two most probable models within the TMNCH-GARCH specification,
Table 3
Simulated and estimated parameter values of the TMNCH-GARCH model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True values</th>
<th>Transformed</th>
<th>Classical estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{11}$</td>
<td>0.2</td>
<td>$\rho_{11}^*$</td>
<td>-1.38</td>
</tr>
<tr>
<td>$\mu_{11}$</td>
<td>-0.32</td>
<td>$\mu_{11}^*$</td>
<td>-0.32</td>
</tr>
<tr>
<td>$\omega_{11}$</td>
<td>0.03</td>
<td>$\omega_{11}^*$</td>
<td>-3.50</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.25</td>
<td>$\alpha_{11}^*$</td>
<td>-1.38</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.85</td>
<td>$\beta_{11}^*$</td>
<td>-0.16</td>
</tr>
<tr>
<td>$\omega_{12}$</td>
<td>0.03</td>
<td>$\omega_{12}^*$</td>
<td>-3.50</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.03</td>
<td>$\alpha_{12}^*$</td>
<td>-3.50</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.94</td>
<td>$\beta_{12}^*$</td>
<td>-0.06</td>
</tr>
<tr>
<td>$\rho_{21}$</td>
<td>0.5</td>
<td>$\rho_{21}^*$</td>
<td>0</td>
</tr>
<tr>
<td>$\mu_{21}$</td>
<td>0.05</td>
<td>$\mu_{21}^*$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\omega_{21}$</td>
<td>0.02</td>
<td>$\omega_{21}^*$</td>
<td>-3.90</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.05</td>
<td>$\alpha_{21}^*$</td>
<td>-2.99</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.75</td>
<td>$\beta_{21}^*$</td>
<td>-0.28</td>
</tr>
<tr>
<td>$\omega_{22}$</td>
<td>0.02</td>
<td>$\omega_{22}^*$</td>
<td>-3.90</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.25</td>
<td>$\alpha_{22}^*$</td>
<td>-1.38</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.75</td>
<td>$\beta_{22}^*$</td>
<td>-0.28</td>
</tr>
<tr>
<td>$m$</td>
<td>0.2</td>
<td>$m$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.2</td>
<td>$\phi$</td>
<td>0.2</td>
</tr>
</tbody>
</table>

$\text{Est}$ denotes estimates and $\text{StE}$ denotes Standard Errors.

Table 4
Summary statistics for the rates of return of the S&P 500 index

| Rates of return $y_{i,t}$ | $| y_{i,t}|$ | $y_{i,t}^2$ |
|---------------------------|------------|------------|
| Index                     | Mean       | Stdev      | Kurtosis   | Skewness   | LB (50) | LB (50) | LB (50) |
| S&P 500                   | 0.0004     | 0.0097     | 7.8157     | -0.3251    | 85.2621  | 2474.1   | 901.21   |

$N = 3000$ observations, $k = 2$ partitions, $\theta_T^{(k)} = (t_1 = 0.3)$. Est denotes estimates and StE denotes Standard Errors.

Together with their respective posterior probabilities, for the different prior specifications. Within the TNM-GARCH specification, the most probable model, with posterior probability ranging from 0.62 to 0.69, consists of $k = 2$ partitions and a threshold value of 0.0001 at the 50th percentile of the series. Within the TMNCH-GARCH specification, the most probable model, with posterior probability above 0.99, is the model with two partitions and threshold value of 0.0001 at the 50th percentile.

Table 7 shows the marginal posterior probabilities of the number of partitions $k$ which is given by

$$p(k | y) = \sum_{\theta_T^{(k)}} p(k, \theta_T^{(k)} | y).$$

It can be seen that there is very strong evidence for two partitions (one threshold) under both specifications (the marginal posterior probability of two partitions is above 0.93).

Now we consider inference conditional on a single model. In order to choose a ‘global’ best model for our data we compare the most probable TNM-GARCH models with the most probable TMNCH-GARCH models. In Table 8 we present the log-marginal likelihoods and the values of the Akaike information criterion (AIC) and the Schwarz criterion (BIC) for the most probable models within both specifications. Although the AIC favours the most probable TMNCH-GARCH model, the log-marginal likelihood and the BIC provide strong evidence for the most probable TNM-GARCH model. Note that, under a discrete uniform prior on the five models of Table 8, the posterior model probabilities are proportional to the respective marginal likelihoods. Furthermore, the BIC is designed to identify the most probable model. Therefore, both of the above criteria indicate that the TNM-GARCH with two partitions is the
Fig. 3. (a) Rates of return of the analysed index S&P 500. (b) Estimated posterior means of the conditional variances based on the MCMC sample of the parameters of the most probable TNM-GARCH model.

Table 5
Set of possible threshold values for the S&P 500 index

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>10th</th>
<th>20th</th>
<th>30th</th>
<th>40th</th>
<th>50th</th>
<th>60th</th>
<th>70th</th>
<th>80th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−0.0105</td>
<td>−0.0058</td>
<td>−0.0029</td>
<td>−0.001</td>
<td>0.0001</td>
<td>0.002</td>
<td>0.0041</td>
<td>0.0069</td>
<td>0.0112</td>
</tr>
</tbody>
</table>

‘global’ most probable model. For this reason, and since the AIC is well known to favour more complex models, (Kass and Raftery, 1995) we choose to work with this ‘global’ most probable model.

In Table 9 we present the estimates of this model’s parameters. In particular, we present estimates and standard errors for the transformed parameters using classical techniques, and the posterior means and standard deviations of the transformed and the original parameters using the Bayesian approach. Fig. 5 shows trace plots of the MCMC samples from the posterior distribution of the model parameters indicating very good mixing properties. Looking at the parameter estimates of Table 9 we can draw the following conclusions. In the first partition, there is a large component with probability 0.951 and small variance of 0.7929 and a very small component with probability 0.049 and large variance of 5.0186. The estimated kurtosis of the innovations in this partition is 5.496 clearly indicating heavy tails. In the second partition, there is a component with probability 0.309 and small variance of 0.1674 and a component with probability 0.691 and variance of 1.372. The estimated kurtosis of the innovations in that partition is 3.929, which is again indicating the presence of heavy tails. Therefore, the proposed TNM-GARCH model explains the leptokurtosis observed in the series and allows for modelling separately the different underlying dynamics in the two partitions.

From Table 9 and the trace plots in Fig. 5 it can be seen that two of the parameters, namely $\mu_1$ and $\phi_2$, are not significant. This might be an indication that a reduced model specification, excluding $\phi_2$ from the parameter set, would be adequate. However, such a model is not included in the set of competing models considered in our paper and such an inclusion is beyond its scope. We also observe that the classical point estimate of the sum $a + \beta$ equals 0.9976, while the respective posterior mean is 0.997, which indicates persistence in volatility. The posterior sample of $a + \beta$ gives probability to non-stationarity of about 0.15 (i.e. 15% of the sample points are equal or greater than 1). The need to impose stationarity conditions in a Bayesian context is not well understood and not broadly accepted (see, also Vrontos et al., 2000). However, stationarity conditions can be easily imposed by simply rejecting the draws of $a$ and/or $\beta$ if the constraint $a + \beta < 1$ is not satisfied.
Fig. 4. Convergence behaviour of the MCMC stochastic search algorithm for the S&P 500 index. Plots of the posterior probabilities of the different threshold models calculated ergodically every 1000 iterations. Four different prior specifications were used: (a) independent noninformative priors, $\gamma = 2$; (b) independent noninformative priors, $\gamma = 3$; (c) unit information priors, $\gamma = 2$; (d) unit information priors, $\gamma = 3$.

Table 6
Posterior probabilities of the most probable TNM-GARCH and TMNCH-GARCH models for the S&P 500 index using different prior specifications

<table>
<thead>
<tr>
<th>Threshold structure</th>
<th>Non informative prior</th>
<th>Unit information prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma = 2$</td>
<td>$\gamma = 3$</td>
</tr>
<tr>
<td>Panel A: TNM-GARCH specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 2, \theta^{(k)}_0 = (0.0001)$</td>
<td>0.620</td>
<td>0.631</td>
</tr>
<tr>
<td>$k = 2, \theta^{(k)}_0 = (-0.0029)$</td>
<td>0.315</td>
<td>0.305</td>
</tr>
<tr>
<td>$k = 1, \theta^{(k)}_0 = (.)$</td>
<td>0.044</td>
<td>0.031</td>
</tr>
<tr>
<td>$k = 3, \theta^{(k)}_0 = (-0.0105, 0.0001)$</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>$k = 3, \theta^{(k)}_0 = (-0.0105, -0.0029)$</td>
<td>0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>$k = 3, \theta^{(k)}_0 = (-0.0058, -0.0029)$</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>Panel B: TMNCH-GARCH specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 2, \theta^{(k)}_0 = (0.0001)$</td>
<td>0.991</td>
<td>0.994</td>
</tr>
<tr>
<td>$k = 1, \theta^{(k)}_0 = (.)$</td>
<td>0.009</td>
<td>0.006</td>
</tr>
</tbody>
</table>

In Fig. 3(b) we show the estimated posterior means of the conditional variances based on the MCMC output for the ‘global’ most probable model. The return series, plotted in Fig. 3(a), indicate periods of high deviations as well as periods with lower deviations; see for example the period after 1998. A major deviation in the return series is immediately reflected in the TNM-GARCH estimates of conditional volatility. The same is true during periods with
Table 7
Posterior probabilities of the number of partitions in TNM-GARCH models and in TMNCH-GARCH models for the S&P 500 index using different prior specifications

<table>
<thead>
<tr>
<th></th>
<th>Non informative prior</th>
<th>Unit information prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma = 2$</td>
<td>$\gamma = 3$</td>
</tr>
<tr>
<td>Panel A: TNM-GARCH specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>0.044</td>
<td>0.031</td>
</tr>
<tr>
<td>$k = 2$</td>
<td>0.935</td>
<td>0.936</td>
</tr>
<tr>
<td>$k = 3$</td>
<td>0.018</td>
<td>0.020</td>
</tr>
<tr>
<td>Panel B: TMNCH-GARCH specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>$k = 2$</td>
<td>0.991</td>
<td>0.994</td>
</tr>
</tbody>
</table>

Table 8
The log-marginal likelihoods, AIC and BIC values for the most probable TNM-GARCH and TMNCH-GARCH models

<table>
<thead>
<tr>
<th></th>
<th>Log-marginal likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: TNM-GARCH specification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_1$: (0.0001)</td>
<td>11 307</td>
<td>−22 141</td>
<td>−22 073</td>
</tr>
<tr>
<td>$t_1$: (−0.0029)</td>
<td>11 306</td>
<td>−22 137</td>
<td>−22 070</td>
</tr>
<tr>
<td>$t_1$: (−)</td>
<td>11 302</td>
<td>−22 107</td>
<td>−22 064</td>
</tr>
<tr>
<td>Panel B: TMNCH-GARCH specification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_1$: (0.0001)</td>
<td>11 305</td>
<td>−22 169</td>
<td>−22 059</td>
</tr>
<tr>
<td>$t_1$: (−)</td>
<td>11 300</td>
<td>−22 109</td>
<td>−21 199</td>
</tr>
</tbody>
</table>

Table 9
Parameter estimates of the most probable TMN-GARCH model for the S&P 500 index

<table>
<thead>
<tr>
<th>Initial parameters</th>
<th>Transformed parameters</th>
<th>Classical estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PM$</td>
<td>$PS$</td>
<td>$PM$</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.951</td>
<td>0.025</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.158</td>
<td>0.032</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>$7.5 \times 10^{-6}$</td>
<td>$2.5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>−0.070</td>
<td>0.035</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.309</td>
<td>0.056</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.122</td>
<td>0.028</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$4.6 \times 10^{-4}$</td>
<td>$2.3 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.015</td>
<td>0.031</td>
</tr>
<tr>
<td>$\omega$</td>
<td>$3.05 \times 10^{-7}$</td>
<td>$1.7 \times 10^{-7}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.043</td>
<td>0.009</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.954</td>
<td>0.009</td>
</tr>
</tbody>
</table>

$PM$ denotes the posterior mean, $PS$ denotes the posterior standard deviation, $Est$ denotes estimates and $StE$ denotes standard errors.

lower deviations in the returns. Even when the periods of low volatility follow periods of extremely high volatilities the proposed specification manages to adjust its conditional volatility estimates.

In financial models, prediction of the future variances is of particular interest. Being able to find the posterior model probabilities, it seems natural to account for model uncertainty in our predictive inferences. Suppose that we are
interested in $h_{t+1}$, the predictive variance at time $t+1$. Then, its predictive density given data $y$ is given by

$$p(h_{t+1}|y) = \sum_{T_k, \theta \in B} p(h_{t+1}|T_k, \theta, y) p(T_k, \theta|y),$$

(7)

where summation is over a subset $B$ of most probable models and $p(T_k, \theta|y)$ are the respective posterior model probabilities normalized within $B$. After implementing the MCMC stochastic search algorithm it is straightforward to obtain an estimate of (7). First, given a threshold model $T_k, \theta$ we use an MCMC algorithm to obtain a sample of $S$ draws $(\theta_{NM}^*, \theta_{G}^*)$, $i = 1, \ldots, S$, from the posterior distribution of the model parameters. Then, a sample from the posterior predictive distribution $p(h_{t+1}|T_k, \theta, y)$ can be easily obtained by calculating the value of $h_{t+1}^{(i)}$ for each sampled draw $(\theta_{NM}^*, \theta_{G}^*)$, $i = 1, \ldots, S$. Given the sample $h_{t+1}^{(1)}, \ldots, h_{t+1}^{(S)}$, the predictive density $p(h_{t+1}|T_k, \theta, y)$ can be estimated using kernel density estimation. Then (7) suggests that an estimate of $p(h_{t+1}|y)$ can be obtained by weighting all estimates of $p(h_{t+1}|T_k, \theta, y)$ with the corresponding $p(T_k, \theta|y)$. To account for model uncertainty in our empirical application we estimate the conditional variances based on the two most probable threshold models obtained by the MCMC stochastic search algorithm. We present in Fig. 6 the posterior predictive density of the one-step-ahead forecast for the variance of the S&P 500 index based on the two most probable TNM-GARCH models (denoted MP1 and MP2), the NM model without threshold and the BMA based on the two most probable models.

5.2. Value at risk

Value at Risk (VaR) is one of the most important measures—tools in managing financial risk. Financial institutions (e.g. Banks) are allowed to calculate capital requirements on a VaR concept by using their own internal risk models.
The risk-based capital requirement depends on the accuracy of the \( \text{VaR} \) model; if the underlying risk is not properly estimated, this may lead to suboptimal capital allocations with consequences on the financial stability of institutions (Manganelli and Engle, 2001). \( \text{VaR} \) refers to a portfolio’s worst outcome that is expected to occur over a predetermined time horizon and at a given confidence level \( \xi \). From a statistical point of view, the \( \text{VaR} \) computation requires the estimation of the \( \xi \) quantile of the profit–loss distribution. For example, for confidence level \( \xi \) and time horizon of one period

\[
\text{VaR}_{t+1|t}(\xi) = F^{-1}(\xi|\Omega_t),
\]

where \( F \) denotes the return’s cumulative distribution and \( \Omega_t \) is the information set up to time \( t \). The above equation implies that

\[
\text{Prob}(y_{t+1} \leq \text{VaR}_{t+1|t}(\xi)) = \xi.
\]

From the above formulation it is obvious that the successful modelling of the return’s distribution is of great importance in order to provide accurate \( \text{VaR} \) estimations. In our study, we use \( \xi = 0.01 \) and 0.05. The first has been selected by the Basel Committee on Banking Supervision, while the second has been adopted by JP Morgan.

In this subsection, we examine the performance of the proposed TNM-GARCH model at computing accurate \( \text{VaR} \) estimates. Our implementation is concerned with the calculation of the \( \text{VaR} \) for the last 800 data points. We use an iterative procedure where the most probable TNM-GARCH models are estimated to predict the one-day-ahead \( \text{VaR}_{t+1|t} \). In particular we provide and compare the one-day-ahead \( \text{VaR}_{t+1|t} \) estimates produced by: (1) the most

Fig. 6. Posterior predictive densities of the one-step-ahead forecast for the variance of the S&P 500 index based on the most probable model, MP1, the second most probable model, MP2, the normal mixture model without threshold (NM) and the Bayesian model average (BMA) based on the two most probable models.
These results indicate that the most probable TNM-GARCH model, and especially the BMA technique improves the simple NM-GARCH model, together with the actual losses of the S&P 500 index. Note that the NM-GARCH model (dotted line) overestimates the VaR estimates by calculating the $\zeta$-quantile of the corresponding predictive distribution $p(y_{t+1} | y)$, $l > 1$. We can use the following sequential procedure to obtain samples from this predictive distribution. For $j = 2, \ldots, l$ the conditional variance $h_{t+j,i}$, given the value of $y_{t+j-1,i}$, can be computed and draws from the predictive distribution $p(y_{t+1} | y)$ can be obtained. See, also Bauwens and Lubrano (1998, 2002) and Ausin and Galeano (2007).

In order to evaluate the performance of these models, we use the method proposed by Kupiec (1995). Let $T_e$ be the total number of exceptions (the actual loss exceeds the VaR estimate) over the out-of-sample period of $P = 800$ data. Then $T_e$ follows a binomial distribution $T_e \sim B(P, \xi)$ and the appropriate likelihood ratio statistic, under the null hypothesis that the total exception rate ($T_e / P$) equals $\xi$ follows asymptotically the $\chi^2(1)$ distribution. In Table 10 we present the total number of exceptions $T_e$ and the corresponding $P$-values based on the Kupiec test for the evaluation of the VaR estimates taken from the most probable TNM-GARCH model, BMA and the simple NM-GARCH model. These results indicate that the most probable TNM-GARCH model, especially the BMA technique improves considerably on the performance of the simple NM-GARCH model; for the VaR $(0.05)$ case, although all models provide accurate VaR estimates at a statistical level of 5%, we see that the most probable TNM-GARCH model and the BMA have larger $P$-values. For the VaR $(0.01)$ case, the null hypothesis is rejected at 5% level of significance for the simple NM-GARCH model, but not for the TNM-GARCH and BMA. Note that for both levels the VaR’s produced by BMA provide the highest $P$-values. We also illustrate in Fig. 7 the VaR estimates produced by the BMA and the NM-GARCH model together with the actual losses of the S&P 500 index. Note that the NM-GARCH model (dotted line) overestimates the VaR $(0.05)$ with respect to the BMA approach, while underestimating the VaR $(0.01)$.}

Next we use another method in order to compare the VaR’s provided by the three methods (most probable TNM-GARCH, BMA and NM-GARCH model). We study the number of exceptions on a rolling sample of 250 data points. This produces 550 exceptions for every model and for each level. As we use 250 data points the best model that provides VaR estimates would be that which has a mean number of exceptions close to 12.5 for $\xi = 5\%$, and a mean close to 2.5 for $\xi = 1\%$. We also calculate the mean absolute percentage error (MAPE). Having a rolling sample of $RS = 250$ data points and denoting with $T_{e,i}^{RS}$, $i = 1, \ldots, 550$ the number of exceptions in the $i$th rolling sample, the MAPE
Fig. 7. (a) VaR_{t+1|t}(0.05) and (b) VaR_{t+1|t}(0.01) estimates produced by the Bayesian model average of the two most probable TNM-GARCH models (solid line) and the simple NM-GARCH model (dotted line) compared with the actual losses of the S&P 500 index (stars).

Table 11
Mean number of exceptions (mean) and mean absolute percentage error (MAPE) using a rolling sample of 250 data points

|                      | VaR_{t+1|t}(0.05) |             | VaR_{t+1|t}(0.01) |             |
|----------------------|-------------------|-------------|-------------------|-------------|
|                      | Mean   | MAPE      | Mean   | MAPE      |
| Simple NM-GARCH model| 13.79  | 0.0165    | 0.41   | 0.084     |
| TNM-GARCH model      | 11.27  | 0.0105    | 0.98   | 0.061     |
| BMA                  | 11.63  | 0.0115    | 1.29   | 0.059     |

is given by

\[
MAPE = \frac{1}{550} \sum_{i=1}^{550} \left| \frac{\tilde{RS}_{e,i} - \tilde{RS}}{RS} \right|.
\]

The model that provides the smallest MAPE is the most preferable. In Table 11, we present the mean number of exceptions and the MAPE for VaR_{t+1|t}(0.05) and VaR_{t+1|t}(0.01) for the most probable TNM-GARCH model, the BMA and the simple NM-GARCH model. For the VaR_{t+1|t}(0.05) case, the smallest MAPE is given by the most probable TNM-GARCH model, while the mean number of exceptions that is close to 12.5 is that of BMA. For the VaR_{t+1|t}(0.01) case, the BMA clearly outperforms the other models.

In summary, we show that BMA, a strategy that make use of the proposed Bayesian stochastic search algorithm and accounts for model uncertainty, outperforms the other models, while the simple NM-GARCH model is not an appropriate model to provide VaR estimates.
6. Discussion

In this study, we proposed a new class of threshold NM models for the analysis and modelling of financial time series. These models account for the nonlinear behaviour of the observed data, since they allow for the existence of different states or partitions, and model the underlying dynamics within each partition in a different way. In particular, we extend two well known mixture normal GARCH models; the NM model of Bai et al. (2003) and the MNCH model of Haas et al. (2004).

We have developed a Bayesian approach for automatic determination of the thresholds as well as the values of these thresholds. A MCMC stochastic search method has been designed which provides posterior model probabilities and takes into account model uncertainty by using BMA.

We have carried out an empirical application of the proposed method to the S&P 500 index using the two threshold normal mixture GARCH specifications and found evidence for two regimes/partitions. Modelling in a different way the underlying dynamics of the series and accounting for model uncertainty by using BMA provide better VaR estimates and increase out-of-sample performance.

Acknowledgements

The authors thank the associate editor and the two referees for their detailed and constructive advice. They would also like to thank Loukia Meligkotsidou for useful comments and discussions.

References


